rotation produced by a conformal mapping usually change throughout the w-plane, the image in the w-plane of boundaries of regions in the z-plane can look very different.

Points z_0 for which $f'(z_0) = 0$ are called **critical points** of the function $f(z)$. It can be seen from the above argument that the conformal nature of a mapping $w = f(z)$ breaks down at a critical point z_0 of $f(z)$, because at such a point the angle between intersecting curves at z_0, and between their image curves at $w_0 = f(z_0)$ are *not* preserved, and in addition the linear and area scale factors vanish at such points. We have proved the following fundamental theorem.

THEOREM 17.1

the fundamental mapping theorem

Conformal mapping Let $f(z)$ be analytic and single valued in a region of the z-plane. Then, at every point z in the region such that $f'(z) \neq 0$, the conformal mapping $w = f(z)$ preserves angles between intersecting curves in the z-plane, and it also preserves the sense of rotation between intersecting directed curves. The linear scale factor involved in the mapping from the z-plane to the w-plane is $\rho(z) = |f'(z)|$ and the area scale factor is $\rho^2(z) = |f'(z)|^2$. ■

The fact that conformal mappings preserve angles between intersecting curves and their sense of rotation leads to the following rule that determines how regions in the z-plane map onto regions in the w-plane. The rule will be used in the examples that follow.

deciding how a region in the z-plane maps onto a region in the w-plane

Rule for determining how a region in the z-plane is mapped onto a corresponding region in the w-plane by a conformal mapping $w = f(z)$

Let a region R in the z-plane be bounded by a continuous and piecewise smooth contour Γ, and let the z-plane be mapped conformally onto the w-plane by $w = f(z)$. Furthermore, let A and B be any two distinct points on Γ and suppose that the region R lies to the left (right) as the boundary Γ is traversed in the direction from A to B. Then if γ is the image of Γ, A' and B' are the images of A and B, and R' is the image of R, the region R' in the w-plane will lie to the left (right) as γ is traversed in the direction from A' to B'.

The preceding rule implies the following simple test for the determination of regions that correspond under a one-one conformal transformation $w = f(z)$. If Z is any test point in a region of interest in the z-plane, then the corresponding region in the w-plane will be the one containing the point $w = f(Z)$.

Before examining some typical examples of conformal transformations, we will prove the important property that the curves $u =$ constant and $v =$ constant in the w-plane are mutually orthogonal at all points other than at the images of the critical points of $w = f(z)$ in the z-plane.

Setting $w = u + iv = f(z)$ and taking the total derivatives of u and v with respect to x gives

$$\frac{du}{dx} = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\frac{dy}{dx} \quad \text{and} \quad \frac{dv}{dx} = \frac{\partial v}{\partial x} + \frac{\partial v}{\partial y}\frac{dy}{dx}.$$

So, along the curves $u =$ constant and $v =$ constant,

$$0 = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\left(\frac{dy}{dx}\right)_{u=const} \quad \text{and} \quad 0 = \frac{\partial v}{\partial x} + \frac{\partial v}{\partial y}\left(\frac{dy}{dx}\right)_{v=const},$$

where $(dy/dx)_{u=const}$ and $(dy/dx)_{v=const}$ are, respectively, the gradients of $u =$ constant and $v =$ constant in the w-plane. Combining these results at an arbitrary point P that is not the image of a critical point of $w = f(z)$ in the z-plane, and writing $(dy/dx)_{u=const,P} = (dy/dx)_{P(u)}$, and $(dy/dx)_{v=const,P} = (dy/dx)_{P(v)}$, we have

$$\left(\frac{dy}{dx}\right)_{P(u)}\left(\frac{dy}{dx}\right)_{P(v)} = \left(-\frac{\partial u}{\partial x}\bigg/\frac{\partial u}{\partial y}\right)_P\left(-\frac{\partial v}{\partial x}\bigg/\frac{\partial v}{\partial y}\right)_P.$$

However, from the Cauchy–Riemann equations the product of these last factors is seen to be -1, showing that

$$\left(\frac{dy}{dx}\right)_{P(u)}\left(\frac{dy}{dx}\right)_{P(v)} = -1.$$

Thus, as P is an arbitrary point in the w-plane at which the product of the gradients of $u =$ constant and $v =$ constant equals -1, it follows directly that the curves $u =$ constant and $v =$ constant are mutually orthogonal except at points that are the images of critical points of $w = f(z)$ in the z-plane. We have proved the next theorem.

THEOREM 17.2

constant values of the real and imaginary parts of $f(z)$ map onto orthogonal trajectories

$u =$ constant and $v =$ constant are orthogonal trajectories If $w = f(z) = u + iv$ is a single-valued analytic function, the families of curves $u =$ constant and $v =$ constant are mutually orthogonal in the w-plane except at the images of the critical points of $f(z)$. ■

(a) The Linear Transformation $w = az + b$

The simplest nontrivial conformal transformation is the **linear transformation**

$$w = az + b \quad \text{with } a \neq 0. \tag{1}$$

As $a \neq 0$ the transformation between the z- and w-planes is one-one, because

$$z = \left(\frac{1}{a}\right)w - \frac{b}{a}, \tag{2}$$

the geometrical properties of the linear mapping

and the transformation is conformal because $f(z) = az + b$ is an analytic function for all z. As $w' = d/dz[az + b] = a \neq 0$, the linear transformation has no critical points. To understand the geometrical interpretation of the linear transformation, notice first that we can write $a = |a|\exp[i\operatorname{Arg} a]$. As a result $w = az + b$ can be regarded as the combination of the three simple transformations,

$$w_1 = |a|z, \quad w_2 = \exp[i\operatorname{Arg} a]w_1, \quad \text{and} \quad w = w_2 + b.$$

The transformation $w_1 = |a|z$ scales z by the real constant factor $|a|$, so although the image in the w_1-plane of the boundary of an arbitrary region in the z-plane experiences neither a translation nor a rotation, it does experience a uniform *magnification* if $|a| > 1$ and uniform *contraction* if $|a| < 1$.

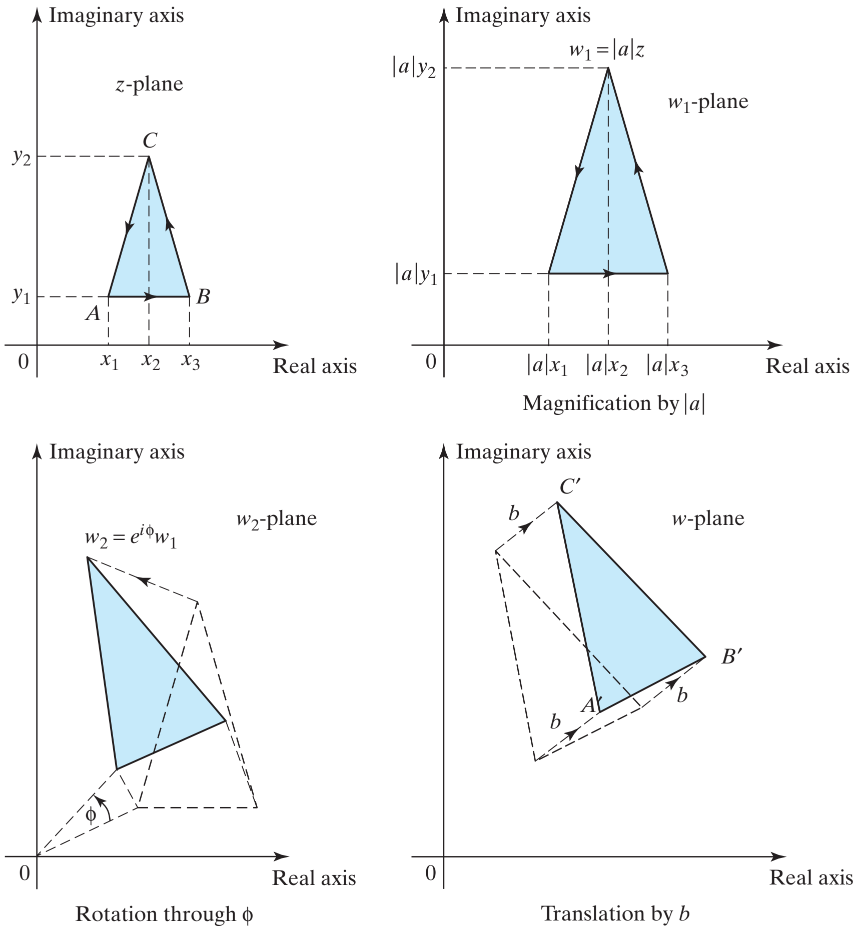


FIGURE 17.3 Successive transformations leading to $w = az + b$.

When complex numbers are multiplied their arguments are added, so setting $\operatorname{Arg} a = \phi$ shows the transformation $w_2 = e^{i\phi}w_1$ produces a uniform *rotation* through an angle ϕ about the origin in the w_1-plane.

Finally, the transformation $w = w_2 + b$ is seen to involve a translation of every point in the w_2-plane by an amount b. So the combined effect of linear transformation (1) on the boundary of any region in the z-plane is to produce first a scaling by a constant factor $|a|$, then a uniform rotation through an angle $\phi = \operatorname{Arg} a$, and finally a uniform translation by an amount b. Thus, a linear transformation *preserves the shapes* of boundaries of regions of interest. The sequence of diagrams in Fig. 17.3 illustrates the typical effect of these successive transformations on a triangular region in the z-plane with its vertices at A, B, and C and the image points A', B', and C' in the w-plane.

To apply the preceding rule to determine the region in the w-plane corresponding to the triangle in the z-plane, we use the fact that the interior of the triangle in the z-plane lies to the left as the boundary is traversed in the direction A, B, and C. Consequently, the corresponding region in the w-plane is the interior of the triangle A', B', and C', because this also lies to the left as the transformed boundary is traversed in the direction A', B', and C'.

It is important always to use a test point with the rule developed earlier in order to check how regions transform. This is because a conformal transformation may map the *interior* of a closed contour Γ in the z-plane onto the *exterior* of its image γ in the w-plane. An example of this type is provided by the inversion mapping that is considered next.

(b) The Inversion Mapping $w = 1/z$

The mapping

$$w = 1/z \tag{3}$$

is called the **inversion mapping**, or sometimes the **reciprocal mapping**. This provides a conformal mapping of the z-plane onto the w-plane, because $f(z) = 1/z$ is a single-valued analytic function with only the simple pole at the origin $z = 0$ where the derivative $w' = -1/z^2$ is not defined. If we set $z = re^{i\theta}$, the mapping becomes

$$w = \left(\frac{1}{r}\right)e^{-i\theta}. \tag{4}$$

This result shows that points on the unit circle $|z| = 1$ map to points on the unit circle $|w| = 1$. However, because of the reversal of the sign of θ, points on the *upper* half of the circle $|z| = 1$ are reflected in the real axis and mapped to points on the *lower* half of the circle $|w| = 1$, and conversely. Furthermore, because $|w| = 1/r$, it follows that points *inside* $|z| = 1$ are mapped to points *outside* $|w| = 1$, and conversely, as shown in Fig. 17.4. This can be confirmed by taking $z = \frac{1}{2}$ as a test point *inside* the unit circle $|z| = 1$, and noticing that it transforms to the point $w = 2$ *outside* the unit circle $|w| = 1$. Notice that the circle in the z-plane and its image in the w-plane are traversed in opposite directions.

The inversion mapping can be regarded as the composition (product) of the two simple transformations

$$Z = \frac{1}{\bar{z}} \quad \text{and} \quad w = \overline{Z}. \tag{5}$$

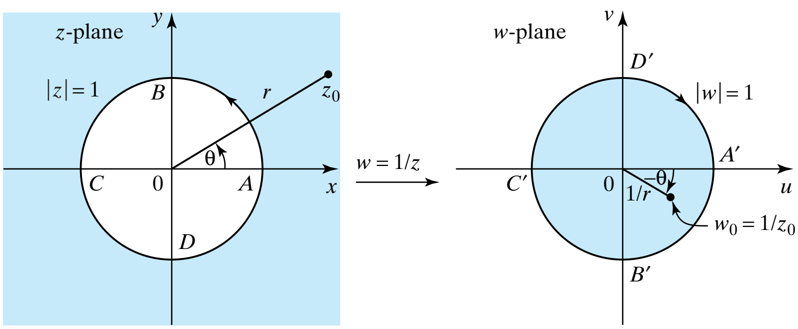


FIGURE 17.4 The inversion mapping $w = 1/z$.

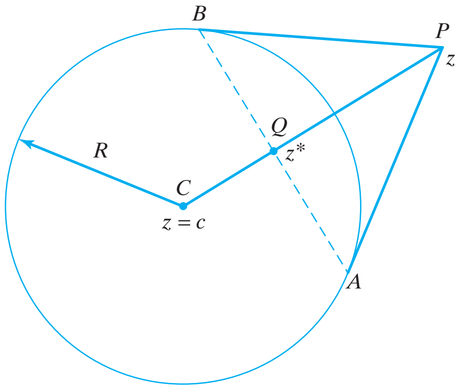


FIGURE 17.5 Inversion in a circle.

To interpret these transformations geometrically we will make use of the general concept of **inversion in a circle**. Consider the circle of radius R in Fig. 17.5, where the point P at z lies outside the circle with its center C at $z = c$, and Q at z^* lies inside it on the radial line CP at its point of intersection with the chord AB drawn from the points A and B where lines from P are tangent to the circle.

the geometrical operation of inversion in a circle

A simple argument using similar triangles shows that

$$|CP| \times |CQ| = R^2,$$

or

$$|z - c|\,|z^* - c| = R^2.$$

The points P and Q in Fig. 17.5 are said to be **symmetric** with respect to the circle with its center at C. Point Q said to be **inverse** to point P and, similarly, point P is inverse to Q. In particular, if $c = 0$ so the circle is centered on the origin, the preceding result implies that points z and z^* that are symmetric with respect to the circle $|z| = R$ are such that

$$z^* = \frac{R^2}{\overline{z}}. \tag{6}$$

Examination of the first transformation in (5) shows that $|Z||\overline{z}| = 1$, so this transformation corresponds to an inversion in the unit circle $|z| = 1$ centered on the origin. The second transformation $w = \overline{Z}$ simply involves the complex conjugate operation, and so can be interpreted as a reflection in the real axis. Thus, the inversion mapping is seen to involve a reflection in the unit circle centered on the origin followed by a reflection in the real axis.

a fixed point of a mapping

A **fixed point** of a mapping f is a point z^* that is left invariant as a result of the mapping, so that $f(z^*) = z^*$. It is easily seen that the inversion mapping has the two fixed points $z = \pm 1$.

The main features of the inversion mapping will become clear if we consider how it maps circles and straight lines. The equation

$$A(x^2 + y^2) + Bx + Cy + D = 0, \tag{7}$$

where the coefficients A, B, C, and D are real, describes a circle of radius $R = (B^2 + C^2 - 4AD)^{1/2}/2|A|$ with its center at $(-B/2A, -C/2A)$ provided $B^2 + C^2 > 4AD$ and $A \neq 0$, and a straight line when $A = 0$. The distance of the center of the circle from the origin is $(B^2 + C^2)^{1/2}/2|A|$, so the circle will *not* pass through the origin if $D \neq 0$, since then $x = 0$, $y = 0$ does not satisfy (7).

If we write $w = u + iv$, the inversion mapping $w = 1/z$ becomes

$$u + iv = \frac{1}{x + iy},$$

from which we find that

$$x = \frac{u}{u^2 + v^2}, \quad y = -\frac{v}{u^2 + v^2}. \tag{8}$$

Substituting (8) into (7) with $A \neq 0$, $D \neq 0$ gives the equation

$$D(u^2 + v^2) + Bu - Cv + A = 0, \tag{9}$$

that describes a circle in the w-plane of radius $\rho = (B^2 + C^2 - 4AD)^{1/2}/2|D|$, with its center at $(-B/2D, C/2D)$. This circle will not pass through the origin in the w-plane if $A \neq 0$, since then $u = v$ does not satisfy (9). Thus, the inversion mapping transforms a circle in the z-plane that does not pass through the origin into a circle in the w-plane that does not pass through the origin.

If, however, $A = 0$ and $D \neq 0$, the straight line in the z-plane given by (7) maps to the circle

$$D(u^2 + v^2) + Bu - Cv = 0 \tag{10}$$

with radius $\rho = (B^2 + C^2)^{1/2}/2|D|$ and its center at $(-B/2D, C/2D)$. As the radius of this circle and the distance of its center from the origin are equal, the circle passes through the origin in the w-plane. Conversely, if $D = 0$ and $A \neq 0$, a straight line in the w-plane will map onto a circle that passes through the origin in the z-plane.

Finally, if $A = D = 0$, the straight line in the z-plane given by (7) will pass through the origin and map onto a straight line in the w-plane that passes through the origin.

In summary, the inversion mapping has the following properties:

summary of the geometrical properties of the inversion mapping

(a) A circle in one plane that does not pass through the origin will map onto a circle in the other plane that does not pass through the origin.

(b) A straight line in one plane that does not pass through the origin will map onto a circle in the other plane that passes through the origin.

(c) A straight line through the origin in one plane will map onto a straight line through the origin in the other plane.

(d) Points inside a unit circle centered on the origin in one plane will map to points outside the unit circle centered on the origin in the other plane, and conversely.

The line $x =$ constant parallel to the imaginary axis is obtained from (7) by setting $A = C = 0$, and examination of the results following (10) shows that this maps onto a circle through the origin in the w-plane with its center on the *real* axis. Similarly, the line $y =$ constant, corresponding to $A = B = 0$ in (10), is seen to map onto a circle through the origin in the w-plane with its center on the *imaginary* axis. Thus, constant coordinate lines map to families of circles through the origin, one with its centers on the real axis and the other with its centers on the imaginary

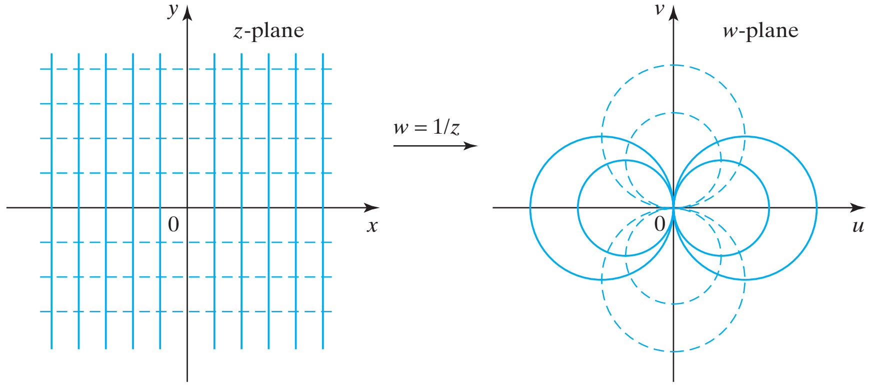


FIGURE 17.6 Mapping of coordinate lines by $w = 1/z$.

axis. As the coordinate lines $x =$ constant and $y =$ constant are orthogonal, the conformal nature of the transformation ensures that the two families of circles are themselves mutually orthogonal, as shown in Fig. 17.6.

The inversion mapping relates directly to the *extended complex plane* introduced at the end of Section 15.3. It will be recalled that the extended complex plane is formed by including in the ordinary complex plane the so-called *point at infinity*, defined as the limit as $R \to \infty$ of all points in the z-plane that lie outside the circle $|z| = R$. As a result, the inversion mapping is seen to map the origin in the z-plane to the point at infinity in the w-plane, and the point at infinity in the z-plane to the origin in the w-plane. If we set $T(z) = 1/z$, the inversion mapping becomes $w = T(z)$, and we can then write $T(0) = \infty$ and $T(\infty) = 0$.

The use of the extended complex plane unifies the treatment of the mapping of straight lines and circles by $w = 1/z$ by allowing straight lines to be regarded as circles of infinite radius.

The effect of an inversion mapping on the square in the z-plane with its sides parallel to the real and imaginary axes shown in diagram (a) on the left of Fig. 17.7 can be seen in the diagram (b) on the right. The sides of the square are seen to map

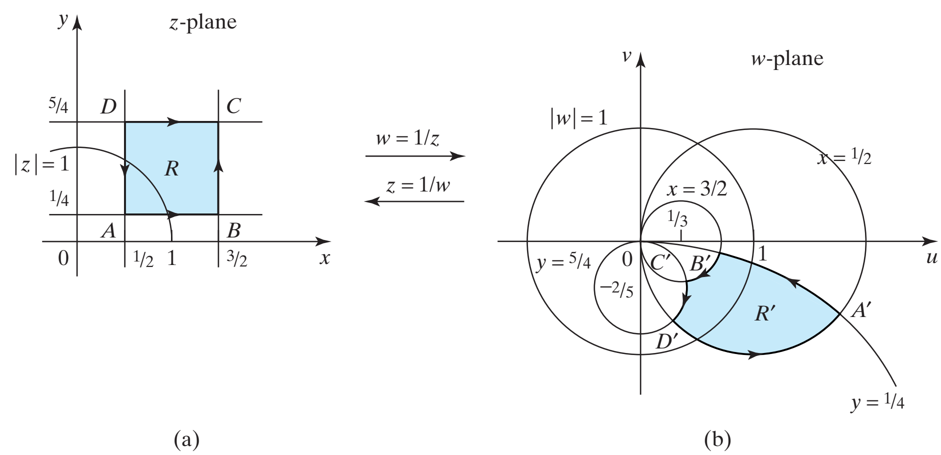


FIGURE 17.7 Inversion mapping of a square.

to four circular arcs, and the rule for determining how regions transform shows that the interior of the square maps to the interior of the region bounded by the circular arcs. For reference purposes the unit circles centered on the origin have been shown in both planes to illustrate how points B, C, and D that lie outside the unit circle in the z-plane map to points inside the unit circle in the w-plane, while point A that lies inside the unit circle in the z-plane maps to a point outside the unit circle in the w-plane. The effect of the reflection in the real axis that is involved in the inverse mapping is also apparent, because a region in the first quadrant in the z-plane has been mapped to a region in the fourth quadrant in the w-plane.

(c) The Linear Fractional Transformation

The transformation

$$w = \frac{az+b}{cz+d}, \tag{11}$$

the linear fractional transformation, or bilinear transformation

is called either the **linear fractional transformation** or the **bilinear transformation**, and sometimes the **Möbius transformation**. It is always possible to assume that $c \neq 0$, because when $c = 0$ the transformation reduces to the linear transformation already considered. Furthermore, we may always assume that $ad - bc \neq 0$, because if $ad - bc = 0$ transformation (11) reduces to a constant.

The inverse mapping

$$z = \frac{b - dw}{cw - a} \tag{12}$$

is also a linear fractional mapping, and as the derivative is

$$w' = \frac{ad - bc}{(cz+d)^2},$$

the mapping is seen to be one-to-one and conformal everywhere with the exception of the point at $z = -d/c$.

Writing the linear fractional transformation in (11) in the form

$$w = \frac{az+b}{cz+d} = \frac{a}{c} + \frac{bc - ad}{c(cz+d)} \tag{13}$$

allows it to be regarded as the sequence of transformations

$$w_1 = cz + d, \quad w_2 = 1/w_1, \quad \text{and} \quad w = (a/c) + \frac{(bc - ad)}{c} w_2. \tag{14}$$

These equations show that a linear fractional transformation can be regarded as the composition of a linear transformation, an inversion mapping, and then another linear transformation.

Having interpreted a general linear fractional transformation in this manner, we can now make use of the general properties of linear transformations and inversion mappings to deduce the general properties of a linear fractional transformation. It is not difficult to see that the transformation (11) maps straight lines and circles onto straight lines and circles, though not necessarily in this order.

summary of geometrical properties of the linear fractional transformation

Furthermore, the definition of symmetry of two points with respect to a circle introduced in (b) earlier when discussing the inversion mapping enables another useful result to be proved: namely, that a pair of points that are symmetric with respect to a circle in the z-plane are mapped by a linear fractional transformation into a pair of points that are symmetric with respect to the image of the circle in the w-plane. The proof of this result is not difficult and so is left as an exercise, but the general result is important because it describes the *symmetry preserving property* of all linear fractional transformations.

When the linear fractional transformation is written in the form

$$w = \frac{(a/c)z + (b/c)}{z + d/c}, \tag{15}$$

it can be seen to be fully determined once the three numbers a/c, b/c, and d/c are specified. We now show how the transformation can be found when three distinct points z_1, z_2, and z_3 that are specified in the z-plane are required to map to three distinct points w_1, w_2, and w_3 that are specified in the w-plane. As three noncollinear points define a circle, it follows that three such points mapping to three other noncollinear points will cause the transformation to map a specific circle in one plane onto a specific circle in the other plane. Similarly, if the three points in one plane are collinear and the three in the other plane are not collinear, the transformation will map a specific straight line in one plane onto a specific circle in the other plane.

Using (11) we can write the difference $w - w_m$ as

$$w - w_m = \frac{(ad - bc)}{(cz + d)(cz_m + d)}(z - z_m), \quad \text{for } m = 1, 2, 3. \tag{16}$$

Forming the differences $w - w_1$, $w - w_2$, $w_3 - w_2$, and $w_3 - w_1$ and combining the resulting expressions leads to the result

a fundamental implicit relationship between w and z

$$\frac{w - w_1}{w - w_2} \cdot \frac{w_3 - w_2}{w_3 - w_1} = \frac{z - z_1}{z - z_2} \cdot \frac{z_3 - z_2}{z_3 - z_1}. \tag{17}$$

This is an *implicit* form of the relationship between w and z that determines the mapping between the specified points in each plane. The *explicit* transformation that produces the required mapping from the z-plane to the w-plane can be obtained from (17) by substituting the numbers z_1, z_2, z_3, w_1, w_2, and w_3 and solving for w in terms of z.

If one of the three points in either plane is the point at infinity, the factors in (17) containing it must be set equal to 1. To understand the reason for this, let us suppose for example that $z_3 = \infty$. Then from (17),

$$\lim_{z_3 \to \infty} \left[\frac{(z - z_1)(z_3 - z_2)}{(z - z_2)(z_3 - z_1)} \right] = \frac{(z - z_1)}{(z - z_2)} \lim_{z_3 \to \infty} \left[\frac{1 - z_2/z_3}{1 - z_1/z_3} \right] = \frac{z - z_1}{z - z_2},$$

using the implicit relationship to find a mapping

confirming that the factors containing z_3 are no longer present and so can be considered to have been set equal to 1. A corresponding result applies if either z_1 or z_2 is the point at infinity, or if any one of w_1, w_2, or w_3 is the point at infinity.

EXAMPLE 17.1

Find the linear fractional transformation that maps the points $z_1 = -1$, $z_2 = 1$, and $z_3 = i$ onto the respective points $w_1 = 0$, $w_2 = 1$, and $w_3 = -i$, and determine how the region R inside the circle through the three points in the z-plane maps onto a region R' in the w-plane.

Solution Substitution into (17) gives

$$\frac{w}{w-1} \cdot \frac{-(1+i)}{-i} = \frac{z+1}{z-1} \cdot \frac{i-1}{i+1},$$

so solving for w shows the required linear fractional transformation to be

$$w = \frac{z+1}{(2+i)z-i}.$$

The circles in the z- and w-planes through the stated points are shown in Fig. 17.8. As the region R *inside* the circle in the z-plane lies to the left as the circle is traversed in the direction z_1, z_2, and z_3, traversing the image points in the w-plane in the order w_1, w_2, and w_3 shows that the image R' of R must lie *outside* the circle in the w-plane. This is easily confirmed by noticing that the point $z = 0$ in R maps to the point $w = i$ in R'. ■

EXAMPLE 17.2

Find the linear fractional transformation that maps the points $z_1 = -1$, $z_2 = 0$, and $z_3 = i$ onto the three points $w_1 = 0$, $w_2 = 1$, and $w_3 = \infty$, and determine how the region R inside the circle through the three points in the z-plane maps onto a region R' in the w-plane.

Solution Substituting z_1, z_2, z_3, w_1, and w_2 into (17), and using the fact that $w_3 = \infty$ enables the factor containing w_3 to be replaced by 1, we find that

$$\frac{w}{w-1} = \frac{z+1}{z} \cdot \frac{i}{i+1}.$$

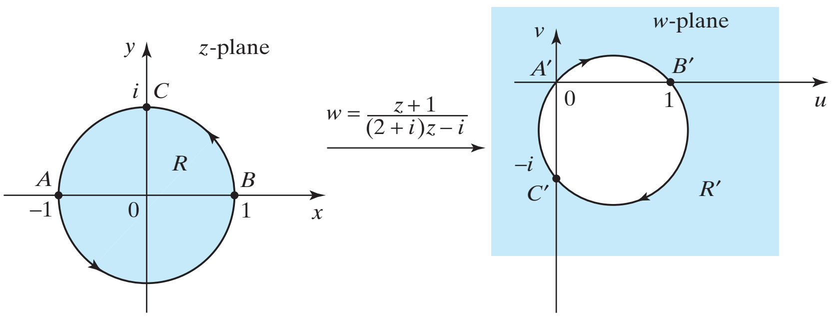


FIGURE 17.8 The mapping $\frac{z+1}{(2+i)z-i}$.

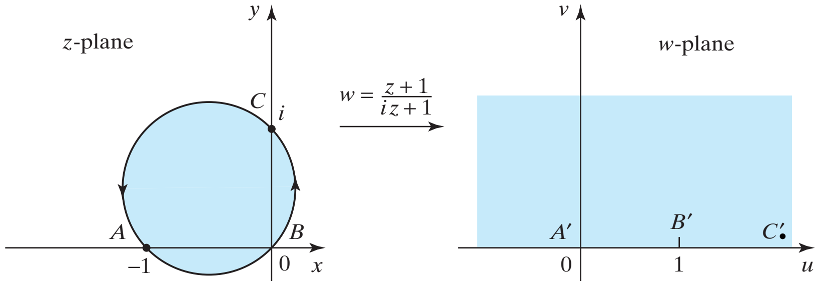


FIGURE 17.9 The mapping $w = \frac{z+1}{iz+1}$.

When solved for w the required linear fractional transformation is found to be given by

$$w = \frac{z+1}{iz+1}.$$

The circle in the z-plane and the corresponding straight line image in the w-plane are shown in Fig. 17.9. The ordering of the points in the two planes shows that as the region R *inside* the circle in the z-plane lies to the left as the circle is traversed in the direction z_1, z_2, and z_3, the image region R' in the w-plane must lie *above* (to the left) as the straight line (real axis) is traversed in the direction w_1, w_2, and w_3 in the w-plane. ■

(d) Mapping Eccentric Circles onto Concentric Circles

how to map eccentric circles onto concentric circles

A linear fractional transformation can map circles onto circles and, when doing so, preserves symmetry. Thus, it can be used to map the region between the eccentric circles in Fig. 17.10a onto the annular region between the concentric circles in Fig. 17.10b.

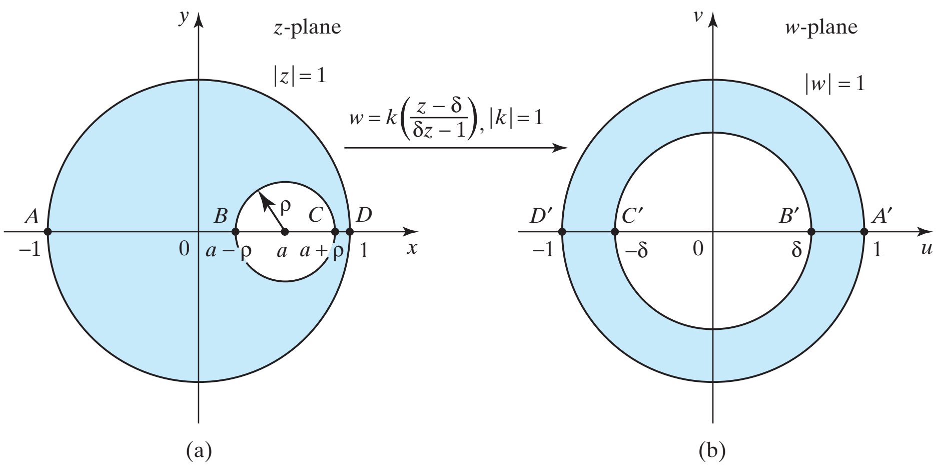


FIGURE 17.10 Mapping eccentric circles onto concentric circles.

To find the required transformation $w = T(z)$, we start from the fact that a linear fractional transformation $T(z)$ can always be written in the form

$$w = T(z) = K\left(\frac{z-\alpha}{z-\beta}\right).$$

So if the center of the inner circle of radius ρ in Fig. 17.10(a) located at $z = a$ is to map to the origin in the w-plane in Fig. 17.10b, we must set $\alpha = a$, so that $T(z)$ becomes

$$T(z) = K\left(\frac{z-a}{z-\beta}\right).$$

The circles in Fig. 17.10(a) are symmetric about the real axis, so this symmetry will be preserved by $T(z)$. In addition, a point z^* that is symmetric relative to $z = a$ with respect to the circle $|z| = 1$ will be mapped onto a point in the w-plane that is symmetric relative to the origin $w = 0$ with respect to the circle $|w| = 1$, so z^* will be mapped to the point at infinity, showing that we must set $\beta = z^*$. The mapping $T(z)$ now takes the form

$$T(z) = K\left(\frac{z-a}{z-z^*}\right).$$

As a and z^* are symmetric with respect to the circle $|z| = 1$, it follows from (6) that $az^* = 1$, but a is real, so $z^* = 1/a$ must also be real. Using this result in $T(z)$ reduces it to

$$w = T(z) = aK\left(\frac{z-a}{az-1}\right).$$

The unit circle $|z| = 1$ maps to the unit circle $|w| = 1$, so recognizing that $|w|^2 = w\overline{w} = 1$ and $z\overline{z} = 1$, and using $w = T(z)$ to form the product $w\overline{w}$, we arrive at the equation

$$1 = w\overline{w} = a^2 K\overline{K}\left(\frac{z-a}{az-1}\right)\left(\frac{\overline{z}-a}{a\overline{z}-1}\right) = a^2 K\overline{K}.$$

This result shows that the factor aK must be of unit modulus, so if k is an arbitrary complex number with unit modulus, $T(z)$ can be written

$$w = T(z) = k\left(\frac{z-a}{az-1}\right), \quad \text{with } |k| = 1.$$

The transformation $T(z)$ maps the circle $|z| = 1$ onto the circle $|w| = 1$, and it preserves symmetry about the real axis in the w-plane. As a is arbitrary, although the image of the inner circle must be symmetric about the real axis in the w-plane, the location of its center will depend on a. The two circles in the w-plane are required to be concentric, so the images of $z_1 = a + \rho$ and $z_2 = a - \rho$ must be symmetric with respect to $w = 0$ at the points $w = \pm\delta$ on the real axis in the w-plane. Thus, $T(z)$ must be such that $T(z_1) = -T(z_2)$, and so

$$\frac{a-\rho-\delta}{\delta(a-\rho)-1} = -\left(\frac{a+\rho-\delta}{\delta(a+\rho)-1}\right).$$

After simplification δ is found to be a solution of the quadratic equation

$$a\delta^2 - (1 + a^2 - \rho^2)\delta + a = 0.$$

Examination of the way the boundaries transform confirms that the region between the eccentric circles in the z-plane maps to the region between the concentric circles in the w-plane.

We have shown that the transformation $w = T(z)$ that maps the region between the eccentric circles in Fig. 17.10a onto the annular region between the concentric circles in Fig. 17.10b is given by

$$w = T(z) = k\left(\frac{z - \delta}{\delta z - 1}\right), \quad |k| = 1, \tag{18}$$

with δ a solution

$$a\delta^2 - (1 + a^2 - \rho^2)\delta + a = 0.$$

(e) The Mapping $w = z^2$

The function

$$w = z^2 \tag{19}$$

how $w = z^2$ maps the z-plane onto the w-plane

is analytic for all z, and so provides a conformal mapping of the z-plane onto the w-plane except at $z = 0$, which is a critical point. Setting $z = re^{i\theta}$ and $w = \rho e^{i\phi}$ in (19) gives $w = r^2 e^{2i\theta} = \rho e^{i\phi}$, so

$$\rho = r^2 \quad \text{and} \quad \phi = 2\theta. \tag{20}$$

Consequently the concentric circles $r = R$ (constant) in the z-plane map onto the concentric circles

$$u^2 + v^2 = R^2$$

in the w-plane, while the radial lines $\theta = \alpha$ (constant) radiating out from the origin in the z-plane map onto the radial lines $\phi = 2\alpha$ in the w-plane.

To make the mapping from the z-plane to the w-plane single valued, it is necessary to restrict θ to any interval of length π. It is usual to restrict z to the upper half of the z-plane so $0 < \theta \leq \pi$ and $r > 0$, because then the upper half of the z-plane maps to the entire w-plane with a cut along the positive real axis, as shown in Fig. 17.11. The image of the region R shown in the z-plane is the region R' in the w-plane. The cut is essential to keep the mapping one-one, because the same transformation also maps the lower half of the z-plane onto the same cut w-plane. Without the cut the function $w = z^2$ maps the entire z-plane *twice* onto the entire w-plane.

Setting $z = x + iy$ and $w = u + iv$ in $w = z^2$ and equating the real and imaginary parts of the equation shows that

$$u = x^2 - y^2 \quad \text{and} \quad v = 2xy. \tag{21}$$

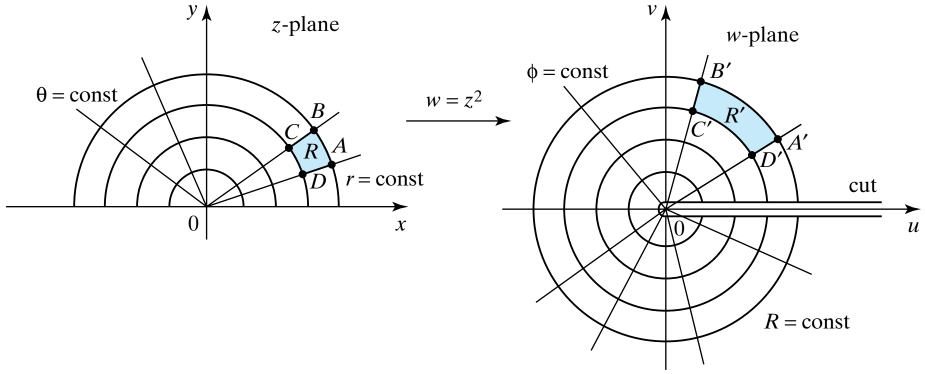


FIGURE 17.11 The mapping $w = z^2$.

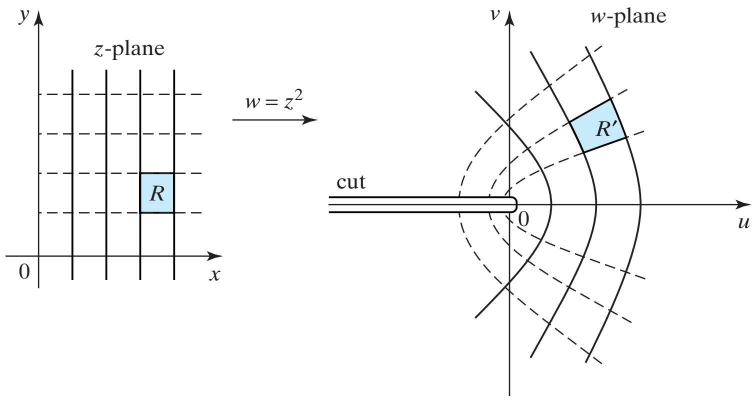


FIGURE 17.12 Mapping of cartesian coordinate lines by $w = z^2$.

So the lines $x = p$ map to the parabolas

$$v^2 = 4p^2(p^2 - u), \tag{22}$$

and the lines $y = q$ map to the parabolas

$$v^2 = 4q^2(u + q^2). \tag{23}$$

This mapping of cartesian coordinate lines in the z-plane onto parabolas in the w-plane is shown in Fig. 17.12, where region R' is the image of region R.

(f) The Function $w = z^{1/2}$

mapping by the branches of the square root function

The **square root function**

$$w = z^{1/2} \tag{24}$$

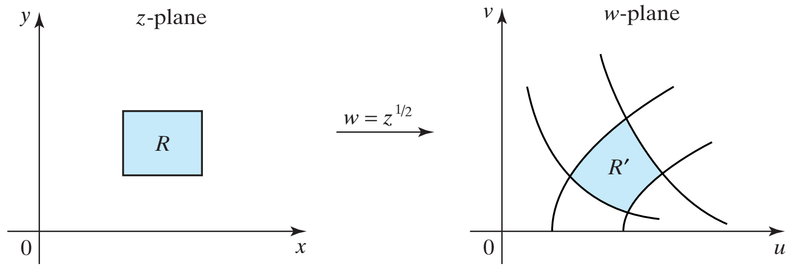


FIGURE 17.13 Mapping of a rectangle in the z-plane by the principal branch of $w = z^{1/2}$.

is the inverse of the mapping considered in (e) above. As the derivative of the square root function is $w' = \frac{1}{2}z^{-1/2}$, the square root function is seen to be an analytic function for all $z \neq 0$, so the conformal nature of the mapping from the z-plane to the w-plane will only fail at the origin. To make the function single valued, we will work with the principal branch of the square root function by setting $z = re^{i\theta}$, and then restricting θ to the interval $-\pi < \theta \leq \pi$, with $r > 0$. If we write $w = u + iv$, the mapping in (24) becomes

$$w = u + iv = r^{1/2}(\cos\theta/2 + i\sin\theta/2), \tag{25}$$

showing that

$$u = r^{1/2}\cos\theta/2 \quad \text{and} \quad v = r^{1/2}\sin\theta/2. \tag{26}$$

If the z-plane is cut along the negative real axis, results (26) show that the principal branch of the square root function maps each point of the cut z-plane once onto the right half of the w-plane, as illustrated in Fig. 17.13. Had the other branch of the square root function been used, where w is determined by

$$w = z^{1/2} = r^{1/2}\left(\cos\frac{(\theta + 2\pi)}{2} + i\sin\frac{(\theta + 2\pi)}{2}\right), \tag{27}$$

each point of the same cut z-plane would have been mapped once onto the left half of the w-plane.

To see how the square root function maps the cartesian coordinate lines in the z-plane onto the w-plane, we set $z = x + iy$ and $w = u + iv$ in (24) and square the result. Equating the real and imaginary parts then shows that

$$x = u^2 - v^2 \quad \text{and} \quad y = 2uv. \tag{28}$$

Thus, the cartesian coordinate lines $x =$ constant and $y =$ constant each map to families of rectangular hyperbolas. The conformal nature of the transformation ensures that the two families of hyperbolas are mutually orthogonal everywhere except at the origin where the critical point of the mapping is located. Figure 17.13 illustrates how the principal branch of the square root function maps a rectangular region in the z-plane onto a curvilinear region in the w-plane.

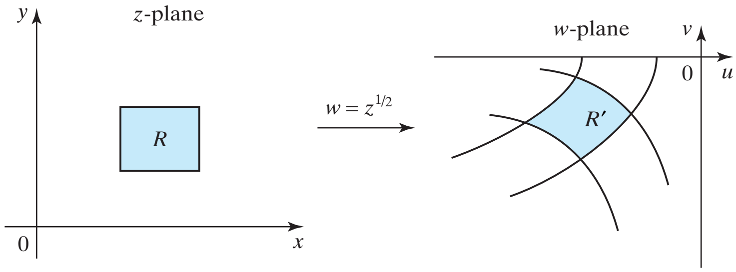


FIGURE 17.14 Mapping of a rectangle in the z-plane by the second branch of $w = z^{1/2}$.

The mapping of the same rectangular region by the second branch of the square root function given in (27) is shown in Fig. 17.14, obtained by rotating the first branch by an angle π.

(g) The Joukowski Transformation $w = z + 1/z$

the Joukowski transformation

The mapping

$$w = z + \frac{1}{z} \tag{29}$$

is called the **Joukowski transformation**, and as $w' = 1 - 1/z^2$ it is seen that w is analytic everywhere except at $z = 0$, and conformal everywhere except at the critical points located at $z = \pm 1$ that map to the points $w = \pm 2$. Setting $z = re^{i\theta}$ in (29), with $-\pi < \theta \leq \pi$ and $w = u + iv$, gives

$$w = u + iv = \left(r + \frac{1}{r}\right)\cos\theta + i\left(r - \frac{1}{r}\right)\sin\theta,$$

so that

$$u = \left(r + \frac{1}{r}\right)\cos\theta \quad \text{and} \quad v = \left(r - \frac{1}{r}\right)\sin\theta. \tag{30}$$

Examination of these results shows that the unit circle $|z| = 1$ maps onto the segment $-2 < u < 2$, $v = 0$, of the real axis in the w-plane, and that its *exterior* maps to the w-plane from which the cut represented by this segment has been removed. The mapping of the z-plane onto the w-plane by the Joukowski transformation is double valued, because the *interior* of the unit circle is also mapped onto this same cut w-plane. The mapping (29) will be single valued if z is restricted to either the interior or the exterior of the unit circle $|z| = 1$.

Setting $z = x + iy$ and $w = u + iv$ in (29) and equating the real and imaginary parts of the equation give

$$u = \frac{x(x^2 + y^2 + 1)}{x^2 + y^2} \quad \text{and} \quad v = \frac{y(x^2 + y^2 - 1)}{x^2 + y^2}. \tag{31}$$

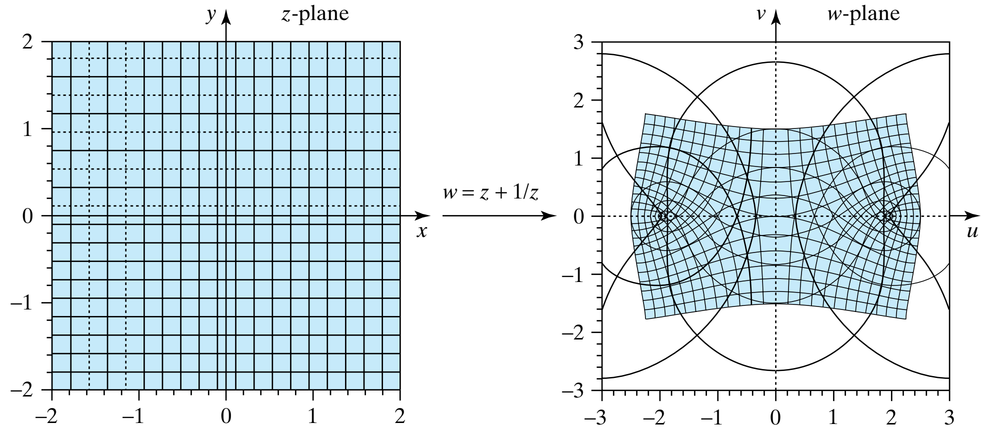


FIGURE 17.15 Mapping of cartesian coordinate lines by $w = z + 1/z$.

These equations determine the way the cartesian coordinate lines $x =$ constant and $y =$ constant map onto the w-plane. Figure 17.15 shows a representative set of mutually orthogonal curves in the w-plane corresponding to a set of cartesian coordinate lines in the z-plane.

Interest in this transformation, which was introduced by the Russian aerodynamicist N. J. Joukowski (1847–1921), first arose because of the way it maps a circle of radius R passing through the point $z = -1$ with its center at a point in the first quadrant of the z-plane onto the w-plane. A typical result of the mapping, called a **Joukowski airfoil profile**, is illustrated in Fig. 17.16. The mapping was used by Joukowski in early studies of the subsonic airflow when calculating the aerodynamic lift of wings with a cross-section in the form of a Joukowski profile.

The inverse mapping from the w-plane to the z-plane is obtained by multiplying the Joukowski transformation in (29) by z and solving the resulting quadratic equation for z in terms of w to obtain

$$z = \frac{1}{2}\left(w + \sqrt{w^2 - 4}\right). \tag{32}$$

The square root function is double valued, so this inverse transformation maps both the *exterior* and *interior* of $|z| = 1$ onto the w-plane, with a cut along the real axis from $w = -2$ to $w = 2$. Because of this it is necessary to use the branch of

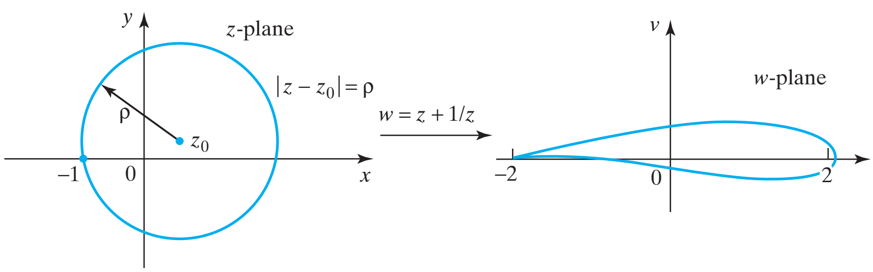


FIGURE 17.16 A typical Joukowski airfoil.

the square root function that is appropriate for the region to be mapped. So, for example, if the exterior of $|z| = 1$ is to be mapped onto the cut w-plane it is necessary to use the branch of the square root function for which

$$|w + \sqrt{w^2 - 4}| > 2.$$

This branch will give a one-one mapping of the upper half of the cut w-plane onto the exterior of the circle $|z| = 1$ in the upper half of the z-plane, with a corresponding mapping of the lower half of the cut w-plane onto the exterior of the circle $|z| = 1$ in the lower half of the z-plane.

(h) The Mappings $w = \sin z$ and Arcsin z

mapping by the sine function and its inverse

The next mapping to be considered is

$$\boxed{w = \sin z} \tag{33}$$

and its inverse Arcsin z.

The function $f(z) = \sin z$ is an entire function, and its critical points are determined by the zeros of $f'(z) = \cos z$ that occur when $z = (k + \frac{1}{2})\pi$ for $k = 0, \pm 1, \pm 2, \ldots$. This means that the mapping $w = \sin z$ will be conformal everywhere except at this infinite set of critical points along the real axis in the z-plane.

Setting $z = x + iy$ and $w = u + iv$ in (33), we have

$$w = \sin z = u + iv = \sin x \cosh y + i \cos x \sinh y,$$

so

$$\boxed{u = \sin x \cosh y \quad \text{and} \quad v = \cos x \sinh y.} \tag{34}$$

As $\sin x$ and $\cos x$ are periodic functions of x, equations (34) show that $w = \sin z$ maps the z-plane infinitely many times onto the w-plane. To make the mapping between the z- and w-planes conformal and one-one, it is necessary to restrict x to lie between any two successive critical points. We choose to require x to lie in the interval $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$ and y to be such that $y \ge 0$, so z lies inside or on the boundary of the semi-infinite strip shown in Fig. 17.17.

As on the side $A_\infty B$ of the semi-infinite strip $x = -\frac{\pi}{2}$ and $y \ge 0$, it follows from (34) that this side must map onto the semi-infinite line segment $A'_\infty B'$ in the w-plane given by $u = -\cosh y$, $y > 0$ and $v = 0$, which lies along the real axis in the w-plane from $-\infty$ to the point $w = -1$. On the line BC, $y = 0$ and $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$,

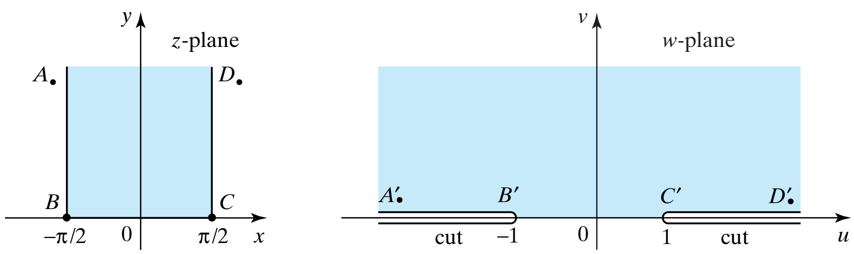


FIGURE 17.17 The mapping of a semi-infinite strip by $w = \sin z$.

so from (34) this line segment is seen to map onto the line segment $B'C'$ given by $-1 \leq u \leq 1$, which is simply the line segment of the real axis in the w-plane extending from $w = -1$ to $w = 1$. Similarly, the side CD_∞ is seen to map to the semi-infinite line segment $C'D'_\infty$ of the real axis in the w-plane extending from $w = 1$ to ∞.

As the interior of the semi-infinite strip lies to the left as the region is traversed in the direction $A_\infty BCD_\infty$, it follows that the interior of the strip must map to the upper half of the w-plane. A similar argument shows that the semi-infinite strip $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$, $y \leq 0$, is mapped by $w = \sin z$ onto the lower half of the w-plane, so that $w = \sin z$ maps the infinite strip $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$ one-one and conformally onto the w-plane cut along the real axis from -1 to $-\infty$ and from 1 to ∞, with the exception of the points $w = \pm 1$ at B' and C' that are the images of the critical points of the mapping located at B and C. These cuts are necessary, because the multivalued nature of $\sin z$ causes the boundaries of each of the semi-infinite strips between successive critical points to map onto the cuts.

The inverse mapping from w to z, denoted by $z = \arcsin w$, is many valued. The mapping can be made one-one by cutting the w-plane along the real axis from $-\frac{\pi}{2}$ to $-\infty$ and from $\frac{\pi}{2}$ to ∞, and then restricting z to any strip of width π that is parallel to the imaginary axis in the z-plane and lies between two adjacent critical points of $\sin z$. When the strip is taken to be $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$, the inverse function is written $z = \operatorname{Arcsin} w$, and this is called the **principal branch of the inverse sine function**.

principal branch of the inverse sine function

If the inverse sine function is considered as a function in its own right, it is usual to interchange w and z and to consider the function $w = \operatorname{Arcsin} z$. The principal branch of the inverse sine function $w = \operatorname{Arcsin} z$ is defined in the z-plane where the cuts along $x < -\frac{\pi}{2}$, $y = 0$, and $x > \frac{\pi}{2}$, $y = 0$, have been made, and $w = \operatorname{Arcsin} z$ is restricted to the strip $-\frac{\pi}{2} \leq \operatorname{Re} w \leq \frac{\pi}{2}$ in the w-plane.

It follows from (34) that the cartesian coordinate lines $x = a$ and $y = b$ map, respectively, to the mutually orthogonal families of hyperbolas and ellipses

$$\frac{u^2}{\sin^2 a} - \frac{v^2}{\cos^2 a} = 1 \quad \text{and} \quad \frac{u^2}{\cosh^2 b} + \frac{v^2}{\sinh^2 b} = 1.$$

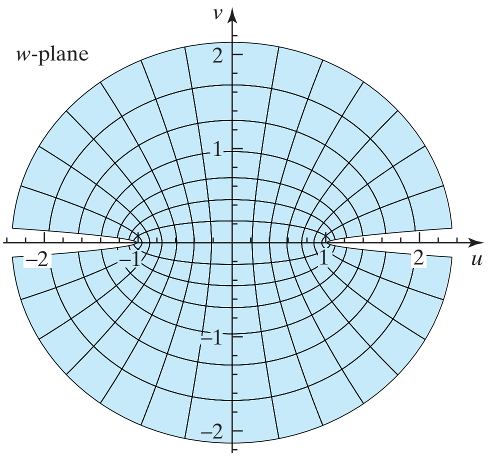


FIGURE 17.18 The mapping of cartesian coordinate lines by $w = \sin z$.

Figure 17.18 illustrates the mapping of these coordinate lines in the z-plane onto the hyperbolas and ellipses in the w-plane by the function $w = \sin z$. The inverse mapping from the w-plane to the z-plane is given by $z = \text{Arcsin}\, w$.

(i) The Mappings $w = \exp z$ and $w = \text{Log}\, z$

The function $\exp z$ is an entire function, so writing it in the form

$$w = \exp(z) = e^x(\cos y + i \sin y) \tag{35}$$

shows that $\exp z$ is periodic in y with period 2π. Thus, $w = \exp z$ will map any strip of width 2π parallel to the imaginary axis one-one and conformally onto the w-plane from which the point $w = 0$ has been deleted. The deletion of the point $w = 0$ is necessary because for no finite z is it true that $\exp z = 0$. The strip $-\pi < y \leq \pi$ is called the **fundamental strip** of the $\exp z$, and from now on y will be restricted to this strip.

the exponential and logarithmic mappings and fundamental strips

Setting $w = u + iv$ in (35) and equating real and imaginary parts give

$$u = e^x \cos y \quad \text{and} \quad v = e^x \sin y. \tag{36}$$

Eliminating y from (36) shows that the cartesian coordinate lines $x = a$ map to the concentric circles $u^2 + v^2 = e^{2a}$. Setting $y = b$ in (36) and eliminating x shows that the cartesian coordinate lines $y = b$ map to the to radial lines (rays) $v = u \tan b$ emanating from the origin. Because of the restriction on y, the strip in the z-plane maps to the w-plane with a cut along the real axis from the origin to $-\infty$, as shown in Fig. 17.19.

In working with the fundamental strip, the inverse function is the principal branch of the logarithmic function $\text{Log}\, w$, and it will provide a one-one and conformal mapping of the w-plane onto the z-plane. If the logarithmic function is considered as a function in its own right, w and z are interchanged and we obtain the function

$$\text{Log}\, z = \ln|z| + i\,\text{Arg}\, z, \quad \text{with } |z| > 0 \text{ and } -\pi < \text{Arg}\, z \leq \pi. \tag{37}$$

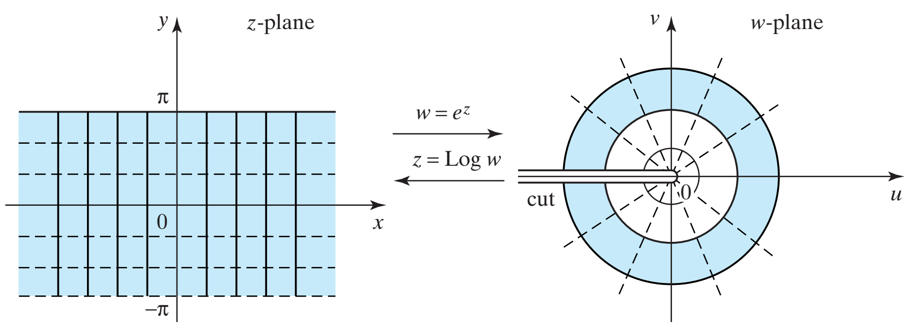


FIGURE 17.19 The mappings $w = \exp z$ and $z = \text{Log}\, w$.

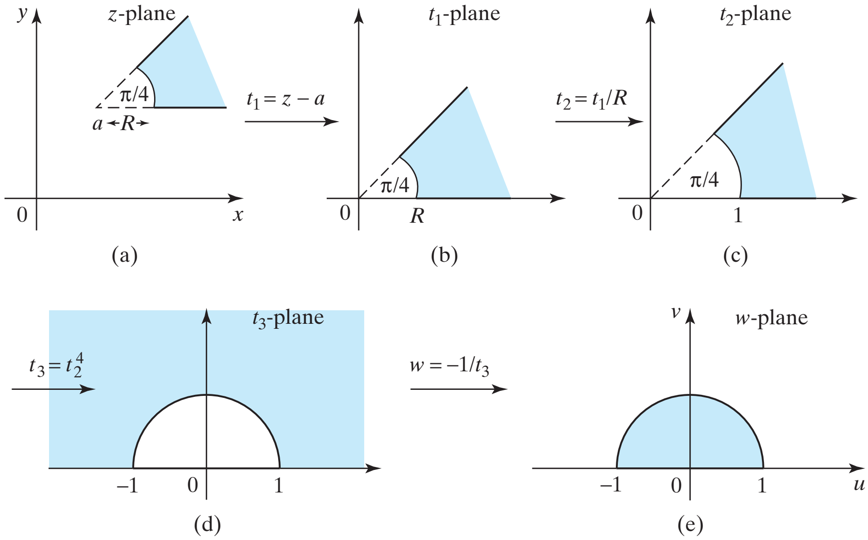


FIGURE 17.20 Mapping an indented semi-infinite sector onto a semicircle.

(j) Composite Mappings

combining mappings to form composite mappings

When considering fundamental mappings such as the inversion mapping and the linear fractional transformation, we have seen how they can be interpreted as a sequence of very simple mappings. The combination of mappings in this manner is called the **composition** of mappings, by analogy with the real variable case where if $w = f(u)$ and $u = g(x)$, the "function of a function" $w = f(g(x))$ is called the composition of the functions g and f.

This approach is also used to build up more complicated mappings when it is required to map a given region onto a more conveniently shaped one. We illustrate this by showing how the interior of the semi-infinite indented wedge-shaped region shown in Fig. 17.20a can be mapped onto the interior of the semicircle $|w| \leq 1$, Im $w \geq 0$, shown in Fig. 17.20e.

The linear mapping $t_1 = z - a$ shifts the vertex of the indented wedge to the origin in Fig. 17.20b without change of scale or rotation. In Fig. 17.20c the mapping $t_2 = t_1/R$ scales the indented wedge so the radius of the circular boundary is 1, again without rotation. In Fig. 17.20d the mapping $t_3 = t_2^4$ opens out the indented wedge so the required region lies in the upper half of the t_3-plane above the unit circle. In Fig. 17.20e the final mapping $w = -1/t_3$ is the inversion mapping, so it maps the indented upper half of the t_3-plane onto the interior of the unit semicircle in the upper half of the w-plane. Eliminating t_1, t_2, and t_3 from these mappings gives the required composite mapping

$$w = \frac{-R^4}{(z-a)^4}.$$

This mapping has a critical point at $z = a$, corresponding to the point $w = \infty$ in the w-plane.

Summary

Conformal mappings have been defined as transformations that preserve both the angle between intersecting curves and the sense of rotation between the curves, when they are mapped from one plane to another. The scale factors determining the stretching of

curves and areas at any point have been derived, and a critical point has been defined as one where the conformal nature of a mapping breaks down. The simple but important linear mapping and its inverse were introduced and their properties combined to give the linear fractional transformation that was then applied to various examples. The quadratic mapping was introduced and shown to map the z-plane twice onto the w-plane and, correspondingly, its inverse mapping by the square-root function was seen to be double valued. The exponential and logarithmic mappings were introduced and composite mappings were defined.

EXERCISES 17.1

1. Describe the effect of the linear transformation $w = 2iz + 3$ when mapping geometrical shapes from the z-plane onto the w-plane. Sketch the image of the rectangle in the z-plane with its corners at $(1, 1)$, $(3, 1)$, $(3, 2)$, and $(1, 2)$, and show the correspondence between corners in the two planes.
2. Describe the effect of the linear transformation $w = (1 + i)z - i$ when mapping geometrical shapes from the z-plane to the w-plane. Sketch (a) the image of the unit circle $|z| = 1$ and (b) the image of the ellipse $(x - 3)^2/9 + y^2/4 = 1$. In each case show how four points on the curve in the z-plane map to the w-plane.
3. Find a linear transformation that maps the triangle with its vertices A, B, and C at points $0, 1 + i$, and $2 - i$ in the z-plane onto the similar triangle with vertices A', B', and C' at $1 - i, 5 - i$, and $3 - 7i$ in the w-plane.
4. Find the linear transformation with the fixed point $2 - i$ that maps $z = -i$ to $w = 2 - 3i$.
5. Find the linear transformation with the fixed point $3 + 2i$ that maps $z = 1$ to $w = -7$.
6. In the following transformations find the fixed point z^* when one exists, the angle of rotation α about z^* that is introduced, and the magnification factor ρ:

$$\text{(a)}\ w = 2z + 1 - 3i. \quad \text{(b)}\ w = iz + 4. \quad \text{(c)}\ w = z + 1 - 2i.$$

7. Find a linear transformation $w = az + b$ that maps the infinite strip $k < y < k + h$ in the z-plane onto the strip $0 < u < 1$ in the w-plane in such a way that $w(ik) = 0$.
8. Find a linear transformation $w = az + b$ that maps the infinite strip $k < x < k + h$ in the z-plane onto the strip $0 < u < 1$ in the w-plane in such a way that $w(k) = 0$.
9. Given that $w = 1/z$, find the image in the w-plane of the family of parallel straight lines $y = x + c$ in the z-plane.
10. By using the symmetry properties of linear fractional mappings, or otherwise, find how $w = z/(z - 1)$ maps the annulus $1 \le |z| \le 2$ in the z-plane onto the w-plane.

In Exercises 11 through 14 find the linear fractional transformation that maps the three given points in the z-plane onto the three given points in the w-plane. Determine the region in the w-plane that corresponds to the region to the left of the given points in the z-plane when the points are traversed in the order z_1, z_2, and z_3.

11. Map points $z_1 = i$, $z_2 = -i$, and $z_3 = 1$ onto the points $w_1 = -1$, $w_2 = 1$, and $w_3 = \infty$.
12. Map the points $z_1 = -1$, $z_2 = -i$, and $z_3 = 1$ onto the points $w_1 = -3 + i$, $w_2 = (2 - 4i)/5$, and $w_3 = 1 + i/3$.
13. Map the points $z_1 = 1$, $z_2 = 2 + i$, and $z_3 = i$ onto the points $w_1 = i$, $w_2 = (-1 + 2i)/5$, and $w_3 = 1/3$.
14. Map the points $z_1 = -1$, $z_2 = 1$, and $z_3 = \infty$ onto the points $w_1 = i$, $w_2 = -i$, and $w_3 = 1$.
15. Prove that the function $w = \exp(\pi z/a)$ maps the infinite strip of width a in the z-plane shown in the diagram on the left of Fig. 17.21 onto the upper half of the w-plane in the manner shown in the diagram on the right. Determine the images in the w-plane of the lines $x = c$ and $y = k$.

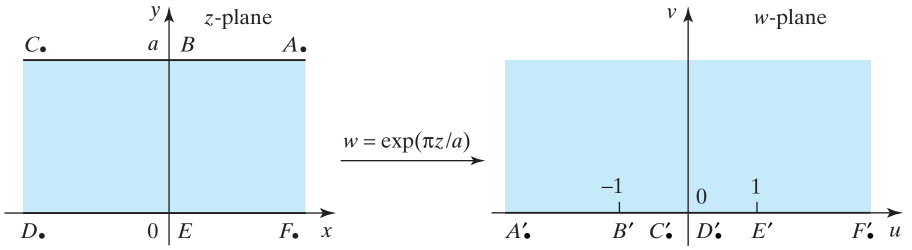


FIGURE 17.21 Mapping by $w = \exp(\pi z/a)$.

16. Prove that the function $w = \sin(\pi z/a)$ maps the semi-infinite strip of width a in the z-plane shown in the diagram on the left of Fig. 17.22 onto the upper half of the w-plane in the manner shown in the diagram on the right. Determine the images in the w-plane of the lines $x = c$ and $y = k$.

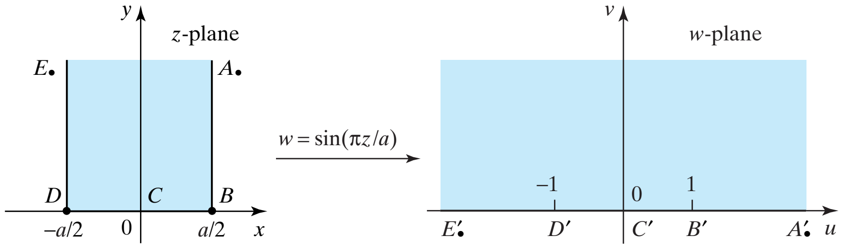


FIGURE 17.22 Mapping by $w = \sin(\pi z/a)$.

17. Prove that the function $w = \cos(\pi z/a)$ maps the semi-infinite strip of width a in the z-plane shown in the diagram on the left of Fig. 17.23 onto the upper-half of the w-plane in the manner shown in the diagram on the right. Determine the images in the w-plane of the lines $x = c$ and $y = k$.

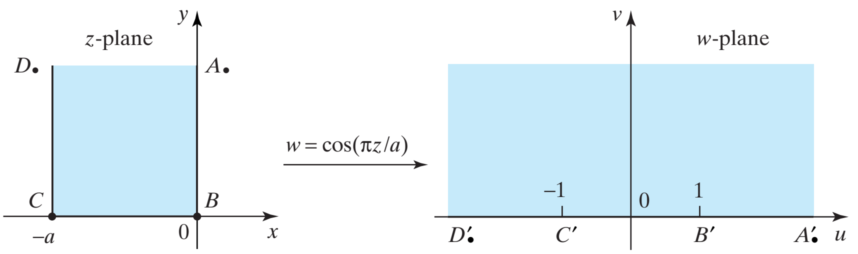


FIGURE 17.23 Mapping by $w = \cos(\pi z/a)$.

18. Prove that the function $w = \cosh(\pi z/a)$ maps the semi-infinite strip of width a in the z-plane shown in the diagram on the left of Fig. 17.24 onto the upper half of the w-plane in the manner shown in the diagram on the right. Determine the images in the w-plane of the lines $x = c$ and $y = k$.

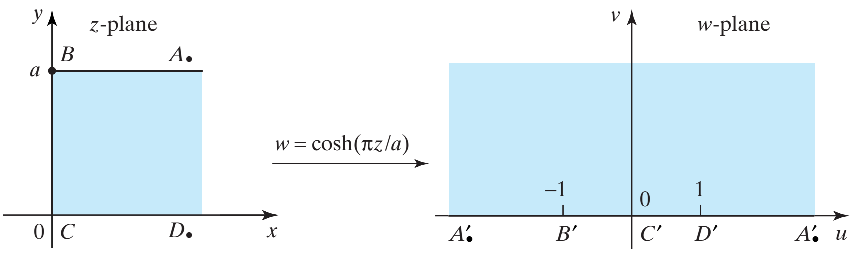


FIGURE 17.24 Mapping by $w = \cosh(\pi z/a)$.

19. Prove that the function $w = (\frac{1+z}{1-z})^2$ maps the interior of the unit semicircle in the z-plane in the diagram on the left of Fig. 17.25 onto the upper half of the w-plane in the manner shown in the diagram on the right.

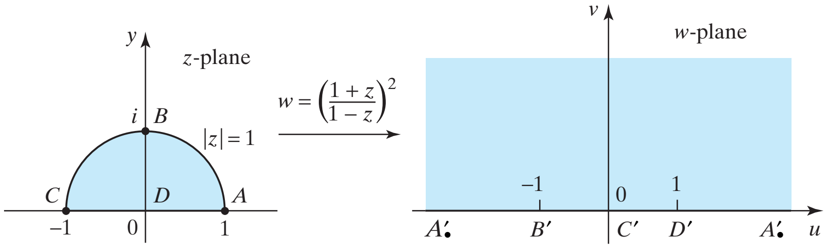


FIGURE 17.25 Mapping by $w = (\frac{1+z}{1-z})^2$.

(Hint: First find the image of $(1+z)/(1-z)$ in the unit circle $|z| = 1$.)

20. Given that $w = z + k/z$, with k real, find the image in the w-plane of the lines $x = c$ and $y = d$. Find the values of k and R such that for given real a and b the transformation will map the circle $|z| = R$ onto the ellipse

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1$$

in the w-plane.

21. Verify that $w = k(\frac{z-z_0}{z-\bar{z}_0})$, with $|k| = 1$ and z_0 an arbitrary point in the upper half of the z-plane, maps the upper half of the z-plane onto $|w| < 1$ and z_0 to the point $w = 0$.

22. Verify that $w = k(\frac{z-z_0}{\bar{z}_0 z-1})$, with $|k| = 1$ and z_0 an arbitrary point such that $|z_0| < 1$, maps $|z| < 1$ onto $|w| < 1$ and z_0 to the point $w = 0$.

23. Show that $w = \tanh z$ maps the semi-infinite strip $0 < y < \pi/2a$ in the diagram on the left of Fig. 17.26 onto the upper half of the w-plane in the manner shown in the diagram on the right.

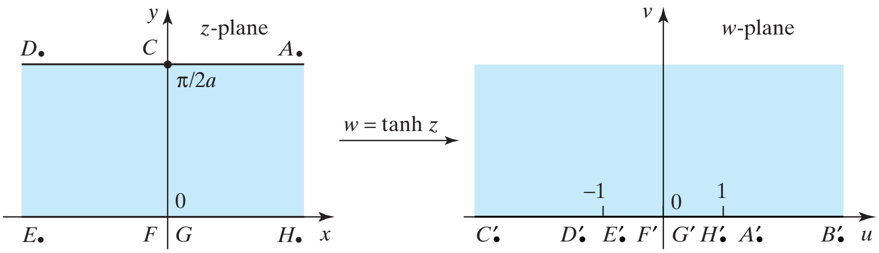


FIGURE 17.26 The mapping $w = \tanh z$.

24. Show that $w = [(1+z^n)/(1-z^n)]^2$ maps the sector in the diagram on the left of Fig. 17.27 onto the upper half of the w-plane in the manner shown in the diagram on the right in the w-plane.

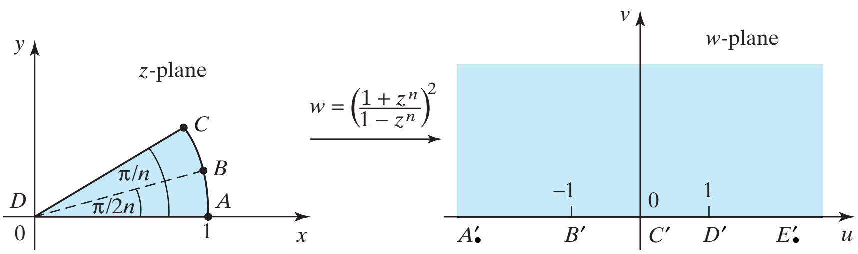


FIGURE 17.27 The mapping $w = [(1+z^n)/(1-z^n)]^2$.

25. Show that $w = (1 - \cos z)/(1 + \cos z)$ maps the semi-infinite strip $0 < x < \pi/2$, $y > 0$ in the diagram on the left of Fig. 17.28 onto the interior of the unit semicircle $|w| = 1$ in the upper half of the w-plane in the manner shown in the diagram on the right.

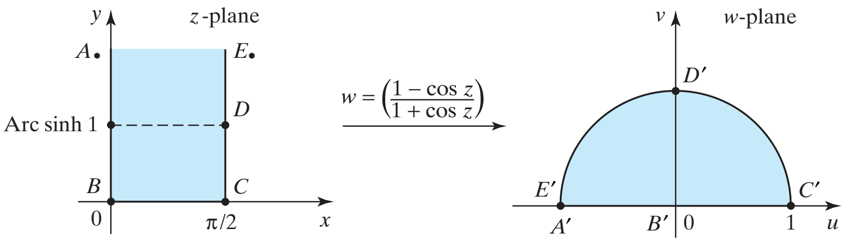


FIGURE 17.28 The mapping $w = (1 - \cos z)/(1 + \cos z)$.

26. Show that $w = \text{Log}(\frac{z-1}{z+1})$ maps the upper half of the z-plane in the diagram on the left of Fig. 17.29 onto the infinite strip $0 < v < \pi$ in the w-plane in the manner shown in the diagram on the right.

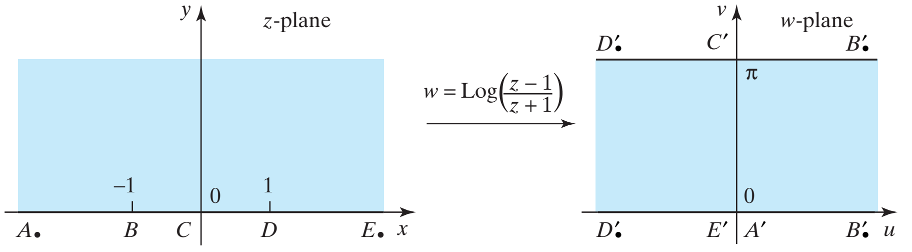


FIGURE 17.29 $w = \text{Log}(\frac{z-1}{z+1})$.

17.2 Conformal Mapping and Boundary Value Problems

boundary value problem for the Laplace equation

The concept of a **boundary value problem** was introduced in connection with the maximum/minimum property of harmonic functions $\phi(x, y)$ (see Theorem 14.17), in which the two independent variables x and y are solutions of the **Laplace equation**

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0. \tag{38}$$

Solutions of Laplace's equation are also called **potential functions** because of the role played by the *gravitational potential* that determines the gravitational force acting on a body and the *electric potential* in space caused by a potential distribution on electrically conducting walls present in, and possibly bounding, the space. In future the Laplace equation will be written $\Delta\phi = 0$ where, as in Chapter 13, the **differential operator** Δ called the **Laplacian operator** in two space dimensions is defined as

$$\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

and $\Delta\phi$ is read "Laplacian ϕ."

In complex analysis only the two-dimensional Laplacian is involved, but in other branches of mathematics both two- and three-dimensional Laplacians occur. To avoid confusion, the two-dimensional Laplacian of ϕ is often denoted by $\Delta_2\phi$ and the three-dimensional Laplacian by $\Delta_3\phi$.

Dirichlet and Neumann boundary conditions

The simplest boundary value problems for the Laplace equation involve specifying either ϕ on the boundary Γ of a region R in which ϕ is harmonic, or the derivative of ϕ normal to the boundary Γ, usually denoted by $\partial\phi/\partial n$. The specification of ϕ on the boundary Γ is called a **Dirichlet boundary condition**, and the requirement that ϕ satisfy both (38) and a Dirichlet boundary condition is called a **Dirichlet boundary value problem** for the harmonic function ϕ. The specification of $\partial\phi/\partial n$ on the boundary Γ of R is called a **Neumann boundary condition**, and the requirement that ϕ satisfy both (38) and a Neumann boundary condition is called a **Neumann boundary value problem** for the harmonic function ϕ. Dirichlet and Neumann boundary value problems are also known as **boundary value problems of the first and second kind**, respectively.

Carl Neumann (1832–1925)
A German mathematician and physicist who in 1868 was appointed Professor of Mathematics at the University of Leipzig. His main contributions were to the study of potential theory and to integral equations.

It is not difficult to show that a Dirichlet boundary value problem for a harmonic function ϕ determines ϕ uniquely at every point of R, and that a Neumann boundary value problem for ϕ determines it uniquely apart from an arbitrary additive constant.

A useful application of conformal mapping is to the solution of two-dimensional boundary value problems for harmonic functions. Various quite different methods of solution exist for such problems, but conformal mapping provides a method that offers valuable geometrical insight into the nature of the solution. The approach comes from the fact that if $w = f(z) = u + iv$ is a single-valued analytic function that maps a region R in the z-plane onto a region R' in the w-plane and $\phi(x, y)$ is harmonic in R, the change of variable from (x, y) to (u, v) transforms $\phi(x, y)$ to a function $\Phi(u, v)$ that is harmonic in R'. Furthermore, either a Dirichlet or a Neumann boundary condition at a point P on the boundary Γ of region R is mapped without change to a point P' on the boundary γ of R', where γ is the image of Γ and P' is the image of P under the mapping $w = f(z)$. In some problems Dirichlet and Neumann boundary conditions apply on different parts of a continuous piecewise smooth boundary Γ, and when this occurs these boundary conditions are transferred to the appropriate parts of the transformed boundary γ. Problems of this type are called **mixed boundary value problems**. In applications to steady state temperature distributions, the temperature satisfies Laplace's equation and a Dirichlet condition on a boundary corresponds to the specification of the temperature on the boundary, whereas the specification of a Neumann condition corresponds to the specification of the temperature gradient across a boundary, and hence the heat flow across the boundary because the heat flow is proportional to the temperature gradient.

mixed boundary value problems

The idea behind a conformal mapping approach to the solution of a boundary value problem for the two-dimensional Laplace equation is to use a conformal transformation $w = f(z)$ to transform a region R in the z-plane with a complicated boundary shape, into a region R' in the w-plane with a more simply shaped boundary. Then, if the solution of the simpler boundary value problem can be found,

the conformal mapping can be used in reverse to transform this simpler solution back into the solution for the more complicated region. As the choice of mapping $w = f(z)$ determines the way in which the boundary of a region R with a simple shape is mapped to a region R' with a more complicated boundary shape, a knowledge of the fundamental mapping properties of elementary functions is necessary when using conformal mapping to solve boundary value problems.

showing that a harmonic function remains harmonic under a conformal mapping

We now give a direct proof that a function $\phi(x, y)$ remains harmonic under the change of variable from (x, y) to (u, v) that transforms $\phi(x, y)$ to $\Phi(u, v)$, where $w = f(z) = u + iv$, and $f(z)$ is a single-valued analytic function. From the chain rule, if $u = u(x, y)$, $v = v(x, y)$ and all functions involved are suitably differentiable,

$$\frac{\partial \phi}{\partial x} = \frac{\partial \Phi}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial \Phi}{\partial v}\frac{\partial v}{\partial x}, \tag{39}$$

and

$$\begin{aligned}\frac{\partial^2 \phi}{\partial x^2} &= \left[\frac{\partial}{\partial x}\left(\frac{\partial \Phi}{\partial u}\right)\right]\left(\frac{\partial u}{\partial x}\right) + \left(\frac{\partial \Phi}{\partial u}\right)\left(\frac{\partial^2 u}{\partial x^2}\right)\\ &\quad + \left[\frac{\partial}{\partial x}\left(\frac{\partial \Phi}{\partial v}\right)\right]\left(\frac{\partial v}{\partial x}\right) + \left(\frac{\partial \Phi}{\partial v}\right)\left(\frac{\partial^2 v}{\partial x^2}\right).\end{aligned} \tag{40}$$

Examination of (39) shows that the differentiation operation $\partial/\partial x$ is related to the differentiation operations $\partial/\partial u$ and $\partial/\partial v$ by

$$\frac{\partial}{\partial x} \equiv \frac{\partial u}{\partial x}\frac{\partial}{\partial u} + \frac{\partial v}{\partial x}\frac{\partial}{\partial v}.$$

Using this result in the terms involving $\frac{\partial}{\partial x}(\frac{\partial \Phi}{\partial u})$ and $\frac{\partial}{\partial x}(\frac{\partial \Phi}{\partial v})$ in (40) changes it to

$$\begin{aligned}\frac{\partial^2 \phi}{\partial x^2} &= \frac{\partial^2 \Phi}{\partial u^2}\left(\frac{\partial u}{\partial x}\right)^2 + \frac{\partial^2 \Phi}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial^2 \Phi}{\partial u \partial v} + \frac{\partial^2 \Phi}{\partial v \partial u}\right)\left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial v}{\partial x}\right)\\ &\quad + \frac{\partial \Phi}{\partial u}\left(\frac{\partial^2 u}{\partial x^2}\right) + \frac{\partial \Phi}{\partial v}\left(\frac{\partial^2 v}{\partial x^2}\right).\end{aligned}$$

A corresponding expression exists for $\frac{\partial^2 \phi}{\partial y^2}$, so combining the two results and using the equality of the mixed derivatives $\frac{\partial^2 \Phi}{\partial u \partial v} = \frac{\partial^2 \Phi}{\partial v \partial u}$, which is justified when Φ is continuous and twice differentiable, leads to the result

$$\begin{aligned}\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} &= \frac{\partial^2 \Phi}{\partial u^2}\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right] + \frac{\partial^2 \Phi}{\partial v^2}\left[\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2\right]\\ &\quad + 2\frac{\partial^2 \Phi}{\partial u \partial v}\left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}\right) + \frac{\partial \Phi}{\partial u}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\partial \Phi}{\partial v}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right).\end{aligned} \tag{41}$$

Examination of (41) shows that the last two terms vanish because u and v are harmonic, while the Cauchy–Riemann equations cause the factor multiplying $\partial^2 \Phi/\partial u \partial v$ to vanish. To simplify the equation further, we now make use of result (21) in Section 13.1 where it was shown that

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

and notice that the Cauchy–Riemann equations allow it to be written in either of the following ways:

$$f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y} \quad \text{or} \quad f'(z) = \frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x}. \tag{42}$$

When the results of (42) are used in the two nonvanishing terms that remain in (41), the equation is seen to reduce to

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = |f'(z)|^2 \left(\frac{\partial^2 \Phi}{\partial u^2} + \frac{\partial^2 \Phi}{\partial v^2}\right) \tag{43}$$

or, equivalently, to $\Delta\phi = |f'(z)|^2 \Delta\Phi$.

This last result shows that if $\phi(x, y)$ is harmonic in the z-plane, then $\Phi(u, v)$ is harmonic in the w-plane, with the exception of points in the w-plane that are images of the critical points of the mapping $w = f(z)$ in the z-plane. We have proved the following important result.

THEOREM 17.3

Harmonic functions remain harmonic under a conformal transformation Let $w = u + iv = f(z)$ be a single-valued analytic function and $\phi(x, y)$ be harmonic in a region R. Then if $\phi(x, y)$ becomes the function $\Phi(u, v)$ under the change of variables $u = u(x, y)$ and $v = v(x, y)$, and R' is the image of R under the transformation, the function $\Phi(u, v)$ is harmonic in R'. ■

To see how the boundary conditions transform, notice first that if P' is the image in the w-plane of a point P on the boundary in the z-plane, then as $\Phi(u, v)$ is simply the function $\phi(x, y)$ expressed in terms of the variables u and v, it follows that $\Phi(P') = \phi(P)$. Also, if $(\partial\phi/\partial n)_P = k(P)$ at a point P on the boundary in the z-plane, then because the mapping is conformal it follows that $(\partial\Phi/\partial n)_{P'}$ will still be normal to the transformed boundary curve in the w-plane at P', so that $(\partial\Phi/\partial n)_{P'} = k(P')$. Thus, Dirichlet and Neumann conditions at P on the boundary in the z-plane are transferred directly to the image of P at P' on the boundary in the w-plane.

solving a fundamental boundary value problem

A fundamental Dirichlet boundary value problem that has many applications involves finding the harmonic function ϕ at an arbitrary point P in the upper half of the (x, y)-plane that satisfies piecewise constant Dirichlet conditions on the x-axis. As the result generalizes in an obvious manner, we will only consider the Dirichlet boundary value problem for the Laplace equation when the solution ϕ is required to assume the three piecewise constant values ϕ_1, ϕ_2, and ϕ_3 on the x-axis. That is, we will solve the Laplace equation

$$\Delta\phi = 0, \quad -\infty < x < \infty,\ y > 0$$

subject to the boundary conditions

$$\begin{aligned} \phi(x, 0) &= \phi_1 \quad \text{for } x < x_1,\ y = 0 \\ \phi(x, 0) &= \phi_2 \quad \text{for } x < x_1 < x_2,\ y = 0 \\ \phi(x, 0) &= \phi_3 \quad \text{for } x > x_2,\ y = 0. \end{aligned} \tag{44}$$

This boundary value problem is illustrated in Fig. 17.30.

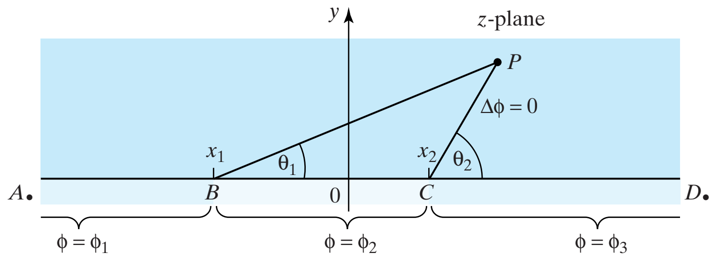


FIGURE 17.30 A piecewise constant Dirichlet boundary value problem.

Inspection shows that the following function ϕ satisfies these boundary conditions:

$$\phi(P) = \phi_3 + \frac{1}{\pi}[(\phi_1 - \phi_2)\theta_1 + (\phi_2 - \phi_3)\theta_2]. \tag{45}$$

To check this it is only necessary to notice that when P in Fig. 17.30 is on the line segment CD_∞ the angles $\theta_1 = \theta_2 = 0$, so $\phi(P) = \phi_3$. Similarly, when P is on the line segment BC, $\theta_1 = 0$ and $\theta_2 = \pi$, so $\phi(P) = \phi_2$, whereas when P is on the line segment $A_\infty B$, $\theta_1 = \theta_2 = \pi$, so $\phi(P) = \phi_1$. The uniqueness of a Dirichlet problem for Laplace's equation then guarantees that (45) is the only solution for this simple boundary value problem, once it has been verified that it is a solution of the Laplace equation.

If Fig. 17.30 is regarded as the complex z-plane, we can write $\theta_1 = \mathrm{Arg}(z - x_1)$ and $\theta_2 = \mathrm{Arg}(z - x_2)$, allowing ϕ to be written

$$\phi(x, y) = \phi_3 + \frac{1}{\pi}(\phi_1 - \phi_2)\mathrm{Arg}(z - x_1) + \frac{1}{\pi}(\phi_2 - \phi_3)\,\mathrm{Arg}(z - x_2).$$

This expression for $\phi(x, y)$ is simply the imaginary part of the complex function

$$w = i\phi_3 + \frac{1}{\pi}(\phi_1 - \phi_2)\,\mathrm{Log}(z - x_1) + \frac{1}{\pi}(\phi_2 - \phi_3)\,\mathrm{Log}(z - x_2).$$

As the function w is analytic for $z \neq x_1, x_2$, its real and imaginary parts are harmonic for $z \neq x_1, x_2$ so, in particular, ϕ must be harmonic for $z \neq x_1, x_2$. The uniqueness of solutions of Dirichlet boundary value problems for harmonic functions then implies that the solution of the boundary value problem in (44) is given by

$$\phi(x, y) = \phi_3 + \frac{1}{\pi}(\phi_1 - \phi_2)\mathrm{Arg}(z - x_1) + \frac{1}{\pi}(\phi_2 - \phi_3)\mathrm{Arg}(z - x_2). \tag{46}$$

Care must be exercised when determining Arg z in terms of the inverse tangent function arctan t. To understand why this is, let point $P(x, y)$ be located at $z = x + iy$ in the upper half of the z-plane, and define θ to be the angle measured counterclockwise from the positive real axis to the line OP drawn from the origin to P, so that $\tan\theta = y/x$. Then, to use (46), an inverse tangent function must be constructed that defines an angle θ that increases *continuously* from 0 to π as P moves counterclockwise around an arc in the upper half of the z-plane, from a point on the positive real axis to one on the negative real axis.

To accomplish this, notice first that the function tan t is defined over the interval $-\pi/2 < t < \pi/2$, and by periodicity elsewhere, so the standard inverse tangent function arctan t cannot be used in (46) when determining θ because it is defined over the wrong interval. However, consideration of the behavior of the function

arctan t over the interval $0 < t < \pi$ shows an Arctan function defined as follows has the required properties:

$$\text{Arctan}\, t = \begin{cases} \arctan t, & t > 0 \\ \pi/2, & t = \pm\infty \\ \pi + \arctan t, & t < 0 \end{cases}. \tag{47}$$

It is this function that must be used in conjunction with (46) when determining ϕ.

The solution of the simplest boundary value problem in which ϕ only assumes two different constant values on the x-axis, with $\phi(x, 0) = \phi_1$ for $x < x_1$, $y = 0$ and $\phi(x, 0) = \phi_2$ for $x > x_1$, $y = 0$, follows directly from the preceding result if we omit the last term (i.e., set $\phi_3 = \phi_2$). If ϕ is required to assume more than three different constant values on the x-axis, result (46) can be extended in an obvious manner. So, for example, if the four constant values ϕ_1, ϕ_2, ϕ_3, and ϕ_4 are involved, and the points separating them on the x-axis are x_1, x_2, and x_3, then in place of (46) we would use

$$\phi(x, y) = \phi_4 + \frac{1}{\pi}(\phi_1 - \phi_2)\,\text{Arg}(z - x_1) + \frac{1}{\pi}(\phi_2 - \phi_3)\text{Arg}(z - x_2) + \frac{1}{\pi}(\phi_3 - \phi_4)\,\text{Arg}(z - x_3).$$

EXAMPLE 17.3

equipotentials

Find the lines of constant electric potential, called either **equipotential lines** or **equipotentials**, in the region between two perpendicular infinitely long electrically conducting walls, when parts of the surfaces are maintained at the constant potentials $\phi_1 = 60$, $\phi_2 = 0$, and $\phi_3 = 20$, as shown in Fig. 17.31.

Solution In space an electric potential ϕ satisfies Laplace's equation so as the conducting walls in Fig. 17.31 are assumed to be infinitely long in the direction perpendicular to the plane of the diagram, and the potentials on the sections of the walls are constant, it follows that ϕ must satisfy the two-dimensional Laplace equation

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = 0.$$

The mapping $w = z^2$ will open up the right angle between the walls in Fig. 17.32a to the half-plane shown in Fig. 17.32b.

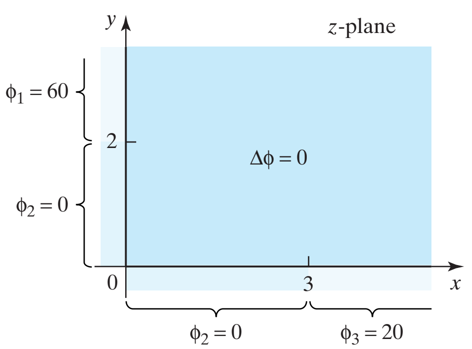


FIGURE 17.31 A Dirichlet problem for the electric potential between two conducting walls.

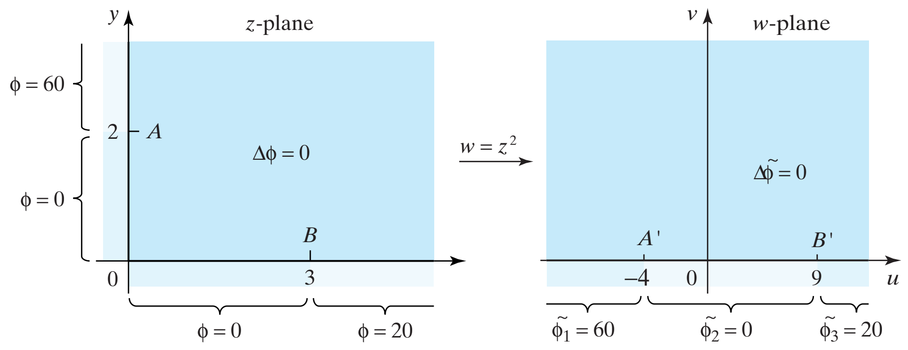


FIGURE 17.32 The effect of the mapping $w = z^2$ on the perpendicular conducting walls.

Setting $w = u + iv$ and and changing from the variables x and y to u and v will cause the potential function $\phi(x, y)$ to become the function $\widetilde{\phi}(u, v)$, and the boundary conditions transform as shown in Fig. 17.32b. The solution of the boundary value problem for $\widetilde{\phi}(u, v)$ follows directly from (46) by replacing z by w, and x_1 and x_2 by $u_1 = -4$ and $u_2 = 9$, respectively, and by setting $\widetilde{\phi}_1 = 60$, $\widetilde{\phi}_2 = 0$, and $\widetilde{\phi}_3 = 20$ to obtain

$$\widetilde{\phi}(u, v) = 60 + \frac{20}{\pi}\mathrm{Arg}(w + 4) - \frac{60}{\pi}\mathrm{Arg}(w - 9).$$

To return to the z-plane we now use the definition of Arctan t in (47), set $z = x + iy$ in $w = z^2$, and write $w = u + iv$ so that $u = x^2 - v^2$ and $v = 2xy$. Then, as $w + 4 = x^2 - y^2 + 4 + i2xy$, we have

$$\mathrm{Arg}(w + 4) = \mathrm{Arctan}\left(\frac{2xy}{x^2 - y^2 + 4}\right)$$

and, similarly,

$$\mathrm{Arg}(w - 9) = \mathrm{Arctan}\left(\frac{2xy}{x^2 - y^2 - 9}\right).$$

So the electric potential at the point (x, y) is seen to be given by

$$\phi(x, y) = 60 + \frac{20}{\pi}\mathrm{Arctan}\left(\frac{2xy}{x^2 - y^2 + 4}\right) - \frac{60}{\pi}\mathrm{Arctan}\left(\frac{2xy}{x^2 - y^2 - 9}\right),$$

for (x, y) in the first quadrant. ■

flux lines

The family of lines $\psi(x, y) =$ constant that form orthogonal trajectories with respect to the equipotentials are called **flux lines**. In electrostatics these are lines of electrostatic force, and in a steady state temperature distribution they correspond to lines of heat flow. If only $\phi(x, y)$ is known, the function $\psi(x, y)$ can be obtained from it by finding the harmonic conjugate function $\psi(x, y)$ using the Cauchy–Riemann equations

$$\frac{\partial\phi}{\partial x} = \frac{\partial\psi}{\partial y} \quad \text{and} \quad \frac{\partial\phi}{\partial y} = -\frac{\partial\psi}{\partial x}.$$

This method is precisely the one given in Section 13.3, by which $\psi(x, y)$ can be recovered from $\phi(x, y)$.

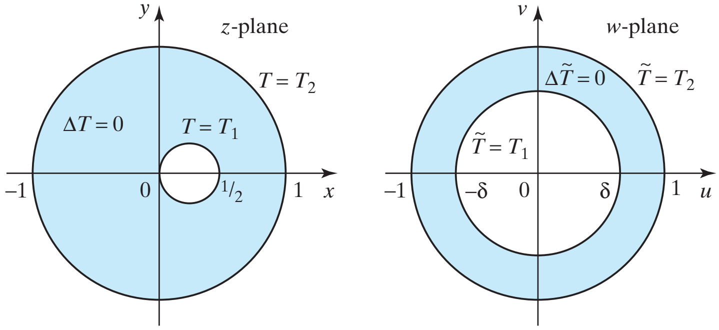


FIGURE 17.33 Equivalent problems in the z-plane and the w-plane.

EXAMPLE 17.4

isothermal lines between eccentric circles

By mapping the region between the eccentric circles on the left of Fig. 17.33 onto the annulus shown on the right, find the lines of constant temperature, called **isothermal lines** or simply **isothermals**, in the region between the eccentric circles when the constant temperature on the inner boundary is T_1 and that on the outer boundary is T_2.

Solution It is shown in Section 18.5 that the two-dimensional steady-state temperature distribution T in a uniform solid is determined by the solution of the two-dimensional Laplace equation $\Delta T = 0$, subject to suitable boundary conditions on the surface of the solid. The two-dimensional formulation of a three-dimensional problem is satisfactory if the solid is in the form of a long uniform bar of constant cross-section and the boundary conditions are constant along the length of the bar, because then the variation of temperature along the length of the bar close to its end faces can be neglected. Under such circumstances the problem reduces to finding the two-dimensional temperature distribution in a lamina in the form of a cross-section of the bar.

When cartesian coordinates are used, the Laplace equation $\Delta T = 0$ satisfied by T is

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0.$$

As T is harmonic, and the problem involves Dirichlet boundary conditions, a conformal transformation $w = f(z)$ with $w = u + iv$ that maps the eccentric circles on the left of Fig. 17.33 onto the concentric circles on the right will lead to an equivalent problem for the temperature $\widetilde{T}$ in the annulus. In what follows the notation $\widetilde{T}(u, v)$ is used to represent $T(x, y)$ after the change of variables from (x, y) to (u, v).

The transformation $w = T(z)$ that maps the eccentric circles onto concentric circles can be found from (18) in Section 17.1. Inspection of the diagram on the left of Fig. 17.10 and a comparison with the geometry of Fig. 17.33 shows that $a = \frac{1}{4}$ and $\rho = \frac{1}{4}$. A simple calculation gives $\delta = 2 - \sqrt{3}$, from which it follows that the required transformation is

$$w = \frac{z - 2 + \sqrt{3}}{(2 - \sqrt{3})z - 1}.$$

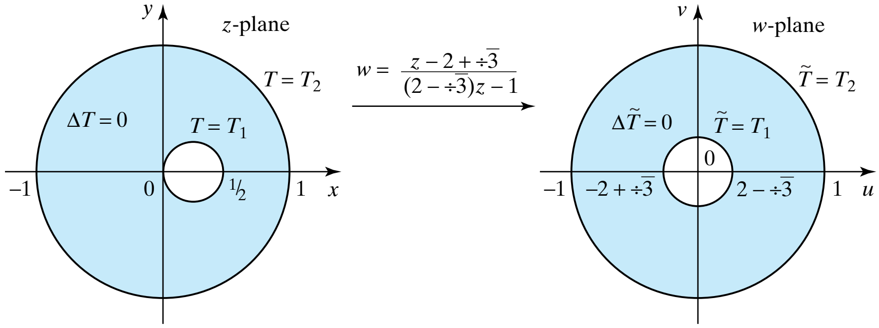


FIGURE 17.34 The mapping $w = \frac{z-2+\sqrt{3}}{(2-\sqrt{3})z-1}$.

The mapping by this function of the region between the eccentric circles onto the annular region is illustrated in Fig. 17.34.

The concentric circular boundaries in the w-plane suggest that $\Delta\widetilde{T} = 0$ should be expressed in terms of cylindrical polar coordinates, leading to the equation

$$\frac{\partial^2 \widetilde{T}}{\partial r^2} + \frac{1}{r}\frac{\partial \widetilde{T}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \widetilde{T}}{\partial \theta^2} = 0.$$

The radial symmetry of the problem in the w-plane shows that the solution must be independent of θ, as a result of which all derivatives with respect to θ vanish, causing Laplace's equation to reduce to the ordinary second order differential equation

$$\frac{d^2 \widetilde{T}}{dr^2} + \frac{1}{r}\frac{d\widetilde{T}}{dr} = 0.$$

Setting $d\widetilde{T}/dr = u$ and integrating gives $u = A/r$, and a further integration then shows the general solution to be

$$\widetilde{T}(r) = A \ln r + B.$$

Matching the integration constants A and B to the boundary conditions $\widetilde{T}(\delta) = T_1$ and $\widetilde{T}(1) = T_2$ gives the solutions in the annulus

$$\widetilde{T}(r) = T_2 - \left(\frac{T_2 - T_1}{\ln(2-\sqrt{3})}\right)\ln r.$$

To return to the (x, y)-plane, it is necessary to express r in terms of x and y, but $r = |w|$, so setting $z = x + iy$ in the expression for w we arrive at the solution

$$T(x, y) = T_2 - \left(\frac{T_2 - T_1}{\ln(2-\sqrt{3})}\right)\ln\left|\frac{x + iy - 2 + \sqrt{3}}{(2-\sqrt{3})(x+iy) - 1}\right|.$$

This solution is complicated, but its typical behavior can be seen by considering the temperature variation along the x-axis, where it reduces to

$$T(x, 0) = T_2 - \left(\frac{T_2 - T_1}{\ln(2-\sqrt{3})}\right)\ln\left|\frac{x - 2 + \sqrt{3}}{(2-\sqrt{3})x - 1}\right|,$$

for $-1 \le x \le 0$ and $1/2 \le x \le 1$. ■

heat flux lines

In Example 17.4 the family of lines $\psi(x, y) =$ constant that form orthogonal trajectories with respect to the isothermals are called **heat flux lines**, and these are lines along which heat flows. When required, the function $\psi(x, y)$ determining the heat flux lines can be obtained from the temperature $T(x, y)$ by finding the harmonic conjugate function $\psi(x, y)$ from the Cauchy–Riemann equations

$$\frac{\partial T}{\partial x} = \frac{\partial \psi}{\partial y} \quad \text{and} \quad \frac{\partial T}{\partial y} = -\frac{\partial \psi}{\partial x},$$

using the method described in Section 13.3.

ideal fluids

Before discussing the next examples it is necessary to preface them with an introduction to the two-dimensional steady flow of an ideal fluid, and its relationship to conformal mapping. An **ideal fluid** is defined as one that is **incompressible**, **inviscid** (free from viscosity), and **irrotational** (its velocity vector $\mathbf{q}$ is such that curl $\mathbf{q} = \mathbf{0}$). The flow of water at low speeds and even of air at subsonic speeds is well approximated by the flow of an ideal fluid.

If in the steady (time-independent) two-dimensional flow of an ideal fluid the velocity vector is $\mathbf{q} = q_1\mathbf{i} + q_2\mathbf{j}$, it is shown in introductory accounts of fluid mechanics that the incompressibility condition follows from the equation of conservation of mass in the form

$$\frac{\partial q_1}{\partial x} + \frac{\partial q_2}{\partial y} = 0 \quad \text{or, equivalently, as div } \mathbf{q} = 0. \tag{48}$$

A simple calculation shows that the irrotational condition curl $\mathbf{q} = \mathbf{0}$ leads to the equation

$$\frac{\partial q_2}{\partial x} - \frac{\partial q_1}{\partial y} = 0, \tag{49}$$

so equations (48) and (49) are seen to take the form of the Cauchy–Riemann equations for the analytic function

$$f(z) = q_1 - iq_2, \tag{50}$$

where the harmonic functions q_1 and q_2 are the components of the fluid velocity vector $\mathbf{q} = q_1\mathbf{i} + q_2\mathbf{j}$.

From vector analysis it is known that if curl $\mathbf{q} = \mathbf{0}$, a scalar function ϕ can always be found with the property that

$$\mathbf{q} = \text{grad}\,\phi, \tag{51}$$

so

$$q_1 = \frac{\partial \phi}{\partial x} \quad \text{and} \quad q_2 = \frac{\partial \phi}{\partial y}. \tag{52}$$

Combining (48) and (52) shows that the real function ϕ satisfies the Laplace equation

$$\Delta\phi = 0, \tag{53}$$

velocity potential, stream function, streamlines, and the complex potential

and hence that ϕ is harmonic. Because of (52) the function ϕ is called the **velocity potential** of the fluid flow. Associated with the velocity potential $\phi(x, y)$ is its harmonic conjugate $\psi(x, y)$, called the **stream function** of the flow, so an analytic function

$$w(z) = \phi(x, y) + i\psi(x, y) \tag{54}$$

can always be defined, called the **complex potential** of the flow, with the property that the curves $\phi(x, y) =$ constant and $\psi(x, y) =$ constant are mutually orthogonal trajectories. The lines along which the stream function is constant are called the **streamlines** of the flow, because the velocity vector is tangent to each point on a streamline. Drawing streamlines enables a flow to be visualized, because any particle of fluid that lies on a streamline will remain on it as it moves steadily across the (x, y)-plane.

We mention here that in many applications the vector $\mathbf{q}$ is often defined in terms of the scalar potential ϕ by writing $\mathbf{q} = -\text{grad}\,\phi$, because it still remains true that curl $\mathbf{q} = \mathbf{0}$. For example, when studying the flow of heat in a steady-state temperature distribution, where ϕ is identified with the temperature T and $\mathbf{q}$ is the heat flow vector, as would be expected the heat then flows in the direction of decreasing temperature. A similar situation also applies in electrostatics.

When required, a stream function can always be found from a given velocity potential $\phi(x, y)$ by the method described in Section 13.3. Result (54) shows that any analytic function can be interpreted as a complex potential, and the streamlines of the flow are then described by the lines along which the stream function is constant. As already mentioned, the functions $\phi(x, y)$ and $\psi(x, y)$ are harmonic conjugates, so the streamlines and lines of constant velocity potential are mutually orthogonal.

Using (52) and (54) together with the fact that ϕ and ψ satisfy the Cauchy–Riemann equations, we can easily show that

$$w'(z) = q_1 - iq_2 \quad \text{and the speed } q = |\mathbf{q}| = \left[\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial y}\right)^2\right]^{1/2}. \tag{55}$$

The connection between the two-dimensional steady flow of an ideal fluid and conformal mapping arises because the complex potential representing the flow in a given region can be mapped conformally onto a different region. This enables the flow in a simple region to be used to determine the flow in a more complicated one.

EXAMPLE 17.5

Interpret the flow of an ideal fluid with the complex potential $w = z^2$, when z is restricted to the first quadrant.

Solution The transformation $w = z^2$ maps the first quadrant in the z-plane onto the upper half of the w-plane. Setting $z = x + iy$ and $w = \phi + i\psi$ and equating real and imaginary parts shows the velocity potential in the w-plane to be $\phi = x^2 - y^2$ and the stream function to be $\psi = 2xy$. The streamlines $\psi =$ constant in the w-plane

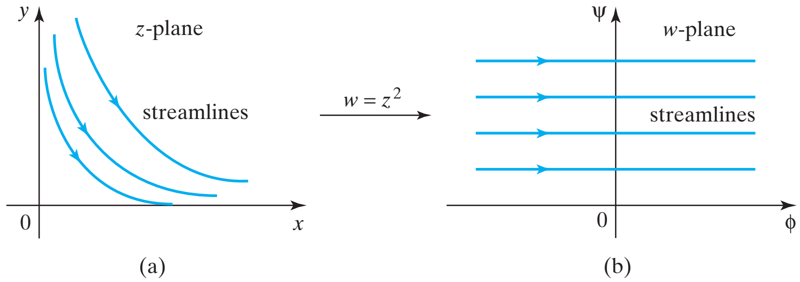


FIGURE 17.35 Flow around two perpendicular walls.

are straight lines parallel to the real axis, so they represent a uniform flow parallel to the real axis as shown in Fig. 17.35b. As no flow crosses the real axis in the w-plane, the axis can be regarded as a rigid wall bounding the flow. The map of this uniform parallel flow in the w-plane onto the z-plane is the family of streamlines $xy =$ constant that form the rectangular hyperbolas shown in Fig. 17.35a. So the complex potential $w = z^2$ describes the flow between two perpendicular walls where, far from the corner, the flow is parallel to a wall.

The velocity components at any point (x, y) in the first quadrant found from from (52) are $q_1 = \partial\phi/\partial x = 2x$ and $q_2 = \partial\phi/\partial y = -2y$, so the flow in the z-plane is in the direction indicated by the arrows in Fig. 17.35a. The speed $q = 2(x^2 + y^2)^{1/2}$ at the point (x, y) follows from (55). It should be recognized that because fluid cannot cross a streamline, in an ideal fluid it is always possible to replace a streamline by a rigid boundary without disturbing the remainder of the flow. ■

EXAMPLE 17.6 Interpret the flow of an ideal fluid with the complex potential

$$w = U\left(z + \frac{1}{z}\right), \quad \text{where } U \text{ is real.}$$

Describe the flow that results when the additional transformation $z = e^{-i\alpha}\zeta$ is made, with α real.

Solution We have seen that the Joukowski transformation maps the exterior of the unit circle $|z| = 1$ in the z-plane onto the w-plane cut along the real axis from $w = -2$ to $w = 2$, as shown in Fig. 17.36.

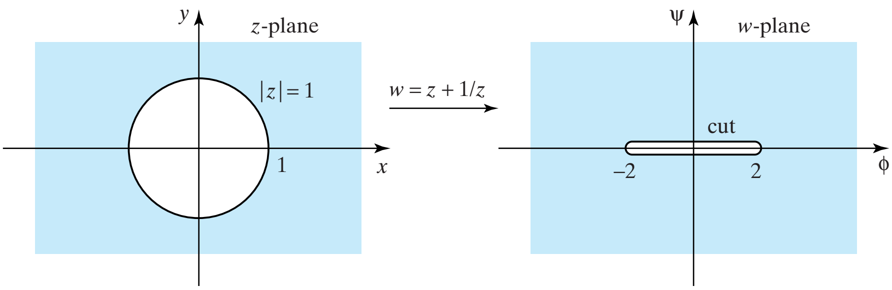


FIGURE 17.36 The effect of the mapping $w = z + 1/z$ on $|z| = 1$.

If we set $w = u + iv$ and $z = x + iy$, routine calculation shows that in cartesian coordinates

$$u = Ux\left(1 + \frac{1}{x^2 + y^2}\right) \quad \text{and} \quad v = Uy\left(1 - \frac{1}{x^2 + y^2}\right),$$

whereas if we set $z = re^{i\theta}$, it follows that in polar coordinates

$$u = (r + 1/r)\cos\theta \quad \text{and} \quad v = (r - 1/r)\sin\theta.$$

When $|x|$ is large the velocity potential $u \approx Ux$, so (52) shows that far from the origin in the z-plane the fluid velocity tends to $\mathbf{q} = U\mathbf{i}$, corresponding to a uniform flow parallel to the x-axis with speed U at infinity. On the unit circle $|z| = 1$ the stream function $\psi = 0$, so this is a *streamline*. Thus, fluid will flow around the unit circle as though it is a solid cylinder of unit radius centered on the origin with its axis perpendicular to the z-plane.

examples of streamlines and equipotentials

The **streamlines** around the unit circle are described by either

$$Uy\left(1 - \frac{1}{x^2 + y^2}\right) = \text{constant} \quad \text{or} \quad \left(r - \frac{1}{r}\right)\sin\theta = \text{constant},$$

whereas the **equipotentials** around the unit circle (lines of constant velocity potential) are described by either

$$Ux\left(1 + \frac{1}{x^2 + y^2}\right) = \text{constant} \quad \text{or} \quad \left(r + \frac{1}{r}\right)\cos\theta = \text{constant}.$$

Figure 17.37 shows some representative streamlines in the z-plane and their images in the w-plane. As no fluid crosses the streamline around the unit circle $|z| = 1$, none will flow across the cut in the w-plane, so the cut can be taken to represent the cross-section of flat plate normal to the z-plane that forms an impenetrable barrier.

The inverse of this transformation can be used to determine the flow past a flat plate when the flow at infinity is incident from the left at an angle α to the plate. From (55) it follows that in the ζ-plane $w_1 = \zeta e^{-i\alpha}$ represents the complex potential of a uniform parallel flow at infinity that is incident from the left at an angle α to the real axis. Consequently, if we use the Joukowski transformation,

$$w = \left(w_1 + \frac{1}{w_1}\right) = \left(\zeta e^{-i\alpha} + \frac{e^{i\alpha}}{\zeta}\right)$$

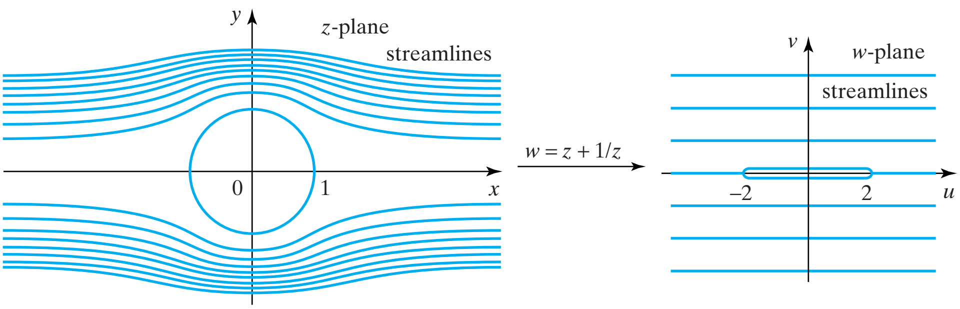


FIGURE 17.37 Flow past a cylinder mapping onto flow parallel to a flat plate.

is the complex potential of a uniform parallel flow that at infinity is incident from the left on the unit circle in the ζ-plane, with the flow at infinity making an angle α with the real axis.

Solving the transformation $\zeta = z + 1/z$ for z, and then interchanging ζ and z, we find that inverse mapping back from the unit circle in the ζ-plane to the z-plane cut from $z = -2$ to $z = 2$ is given by $\zeta = \frac{1}{2}(z + \sqrt{z^2 - 1})$. If we substitute for ζ in the previous result, the required complex potential in the z-plane for flow with speed U past a flat plate formed by the cut from $z = -2$ to $z = 2$, when the flow is incident from the left of the plate and at an angle α, is seen to be given by

$$w = U\left(\frac{1}{2}e^{-i\alpha}(z + \sqrt{z^2 - 4}) + \frac{2e^{i\alpha}}{z + \sqrt{z^2 - 4}}\right).$$

When simplified using the result $z + \sqrt{z^2 - 4} = \frac{1}{4}(z - \sqrt{z^2 - 4})$, this reduces to

$$w = U(z\cos\alpha - i\sqrt{z^2 - 4}\sin\alpha).$$

stagnation point in a flow

In this complex potential, as the square root function has a branch point, we must interpret the square root as $\sqrt{z^2 - 4} = |z^2 - 4|e^{(i/2)(\theta_1+\theta_2)}$, where $z - 2 = |z - 2|e^{i\theta_1}$ and $z + 2 = |z + 2|e^{i\theta_2}$, with $0 \le \theta_1 \le 2\pi$ and $0 \le \theta_2 \le 2\pi$ measured as shown in the cut plane in Fig. 17.38c.

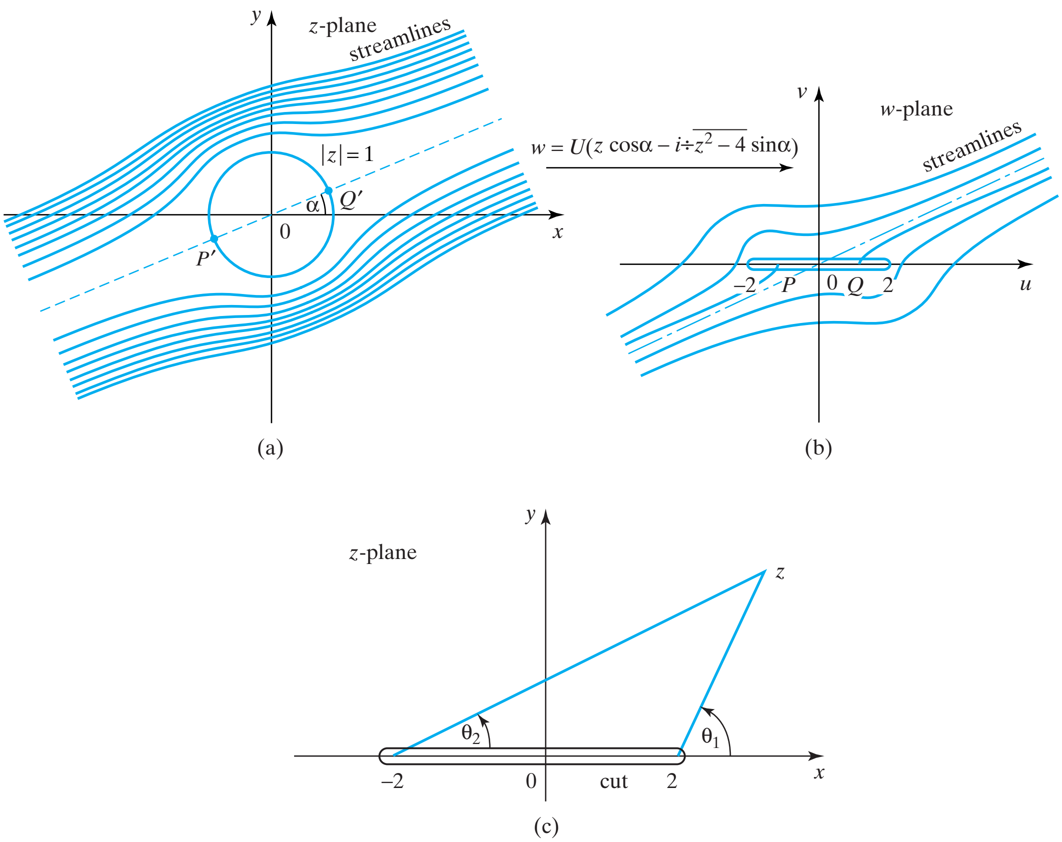


FIGURE 17.38 Inclined flow past a flat plate.

Representative streamlines around the unit circle in the z-plane with the flow at infinity inclined to the x-axis at an angle α are shown in Fig. 17.38a, where the points P' and Q' are streamlines that terminate on the unit circle. These points are called **stagnation points**, because the fluid velocity is zero at such points. Fig. 17.38b shows the inverse mapping of this flow, corresponding to inclined flow around a flat plate in the w-plane, where the stagnation points at P and Q on the plate are the images of the stagnation points P' and Q' in Fig. 17.38b.

The pressure p at any point on a streamline can be found from a result called the **Bernoulli equation**, and for the steady two-dimensional flow of an ideal fluid this takes the form

$$\frac{1}{2}\rho\left(q_1^2 + q_2^2\right) + p = \text{constant},$$

where ρ is the density of the fluid. This shows that the pressures in the vicinity of the stagnation points on either side of the plate apply turning moments to the plate that both act in the same sense. When the plate is broadside to the flow, points P and Q are opposite one another at the center of the plate, about which the flow is then symmetrical. Such a flow provides a good approximation to the actual flow of fluid past a flat plate, and it only fails at the ends of the plate where in the real world the speed of flow is finite, whereas in an ideal fluid it is infinite. The existence of a turning moment about the center line of the plate, which vanishes when the plate is perpendicular to the flow, explains why a boat allowed to drift from rest down a stream will always turn broadside to the direction of flow. ■

The Laplace equation arises in many other steady-state physical situations, the most important of which are in the description of gravitational fields, diffusion, electric current flow, magnetism, and elasticity. When restricted to two space dimensions the real and imaginary parts of an analytic function $w = \phi + i\psi$ can be interpreted as follows:

Application of Laplace's Equation	$\phi(x, y) = $ Constant	$\psi(x, y) = $ Constant
Gravitational fields	Gravitational equipotentials	Lines of force
Diffusion phenomena	Concentration	Lines of flow
Electric current flow	Potential	Lines of current flow
Magnetism	Magnetic potential	Lines of force
Elasticity	Strain function	Stress lines

The development of conformal transformations together with various applications is to be found in references [6.1], [6.2], [6.4], [6.6], [6.8], and [6.9]. A systematic application of conformal transformations is made to hydrodynamics in reference [6.5].

Summary

The Laplace equation is fundamental to the study of heat flow, electricity and magnetism, fluid mechanics, gravitational fields, and elsewhere. This section has shown how conformal mappings can be used to solve certain types of boundary value problems for the Laplace equation in complicated two-dimensional regions bounded by arcs and straight lines. The technique involved first solving a boundary value problem in a simply shaped region bounded by coordinate lines in one plane, and then mapping the region onto one in another plane with a more complicated shape that is of interest. The approach was seen to work because conformal mappings transform harmonic functions in one plane into

harmonic functions in another plane, while the boundary conditions are mapped without change onto the corresponding boundaries. Consequently, the solution of a simple boundary value problem in one plane can be transformed into the solution of a corresponding boundary value problem in a region of more complicated shape in another plane. Applications to various boundary value problems of physical interest were made, including ones to the flow of ideal fluids.

EXERCISES 17.2

1. Let the function $\phi(x, y)$ be harmonic in some region of the (x, y)-plane. If $\phi(x, y)$ becomes $\Phi(u, v)$ under the change of variable $u = x^2 - y^2$ and $v = 2xy$, confirm by direct calculation that the transformation $w = z^2$ leaves Φ harmonic.
2. Using the definition of Arctan t in (47) and setting $t = y/x$, confirm that if (a) P is the point $(\sqrt{3}, 1)$ then Arctan $t = \frac{\pi}{6}$, (b) P is the point $(-2, 2)$ then Arctan $t = \frac{3\pi}{4}$, and (c) if P is the point $(\pm\varepsilon, 2)$, then $\lim_{\varepsilon\to 0}$ Arctan $t = \frac{\pi}{2}$. Find Arctan t when (d) P is the point (4, 1) and (e) when P is the point $(-3, 2)$.
3. Derive the function $\phi(x, y)$ that is harmonic in the upper half of the (x, y)-plane and satisfies the piecewise constant Dirichlet boundary value problem

$$\begin{aligned}
\phi &= \phi_1 \quad \text{on } x < x_1,\ y = 0\\
\phi &= \phi_2 \quad \text{on } x_1 < x < x_2,\ y = 0\\
\phi &= \phi_3 \quad \text{on } x_2 < x < x_3,\ y = 0\\
\phi &= \phi_4 \quad \text{on } x > x_3,\ y = 0.
\end{aligned}$$

4. Derive the function $\phi(x, y)$ that is harmonic in the right half of the (x, y)-plane and satisfies the piecewise constant Dirichlet boundary value problem

$$\begin{aligned}
\phi &= \phi_1 \quad \text{on } y > y_1,\ x = 0\\
\phi &= \phi_2 \quad \text{on } y_2 < y < y_1,\ x = 0\\
\phi &= \phi_3 \quad \text{on } y < y_2,\ x = 0.
\end{aligned}$$

   Is there a simple way of finding $\phi(x, y)$ from (46)?
5. Prove that the transformation $w = (\frac{1+z}{1-z})^2$ maps the interior of the semicircle of radius 1 on the left of Fig. 17.39 onto a half-plane in the manner shown in the diagram on the right. If the semicircle represents a cross-section of a long heat-conducting bar, find the temperature distribution and the isothermals in a cross-section of the bar when the flat boundary AB is maintained at the constant temperature $T = 30$ and the semicircular boundary ACB is maintained at the constant temperature $T = 150$.

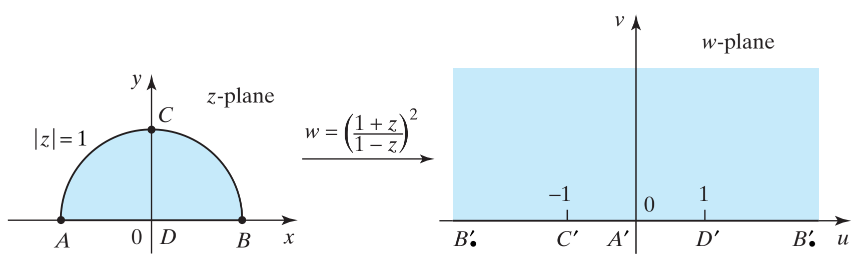


FIGURE 17.39 The mapping $w = (\frac{1+z}{1-z})^2$.

6. Repeat Exercise 5 assuming that the semicircle on the left represents a cross-section of an electrically conducting wall of a cavity. Find the electric potential inside the cavity and the

equipotentials when the flat section of the wall AO is maintained at the constant electric potential $\phi = 20$, the flat section of the wall OB at the constant electric potential $\phi = 100$, and the curved wall ACB at the constant electric potential $\phi = 50$.

7. Prove that the transformation $w = i\left(\frac{1-z}{1+z}\right)$ maps the inside of the circle on the left in Fig. 17.40 onto the upper half-plane in the manner shown in the diagram on the right. If the circle is considered to be the electrically conducting wall of a cavity, find the electric potential and electric force lines inside a cross-section of the cavity if the upper semicircular boundary ABC is maintained at the constant electric potential $\phi = 320$ and the lower semicircular boundary CDA at the constant electric potential $\phi = 100$.

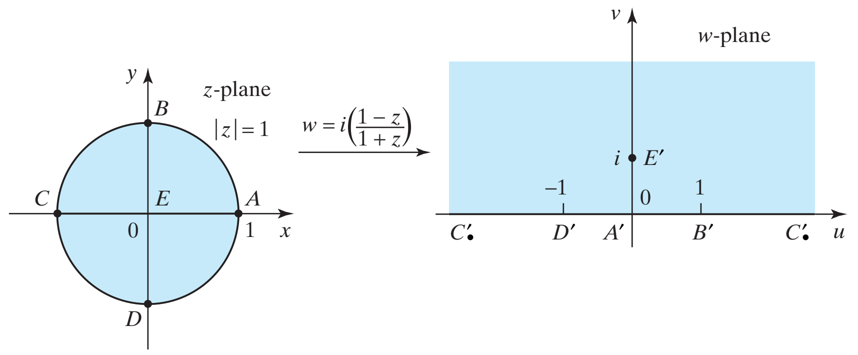


FIGURE 17.40 The mapping $w = i\left(\frac{1-z}{1+z}\right)$.

8. Repeat Exercise 7 assuming the circle to be the cross-section of a long solid heat-conducting cylinder. Find the temperature distribution and the isothermals in a cross-section of the cylinder if the circular boundary CD is maintained at a temperature $T = 50$, the circular boundary DAB is maintained at a constant temperature $T = 200$, and the circular boundary BC is maintained at a constant temperature $T = 0$.

9. Explain why $w = U(z^3 + \frac{1}{z^3})$ is the complex potential of the flow inside the indented wedge shown in Fig. 17.41, in which the flow moves parallel to each wall at infinity with speed U.

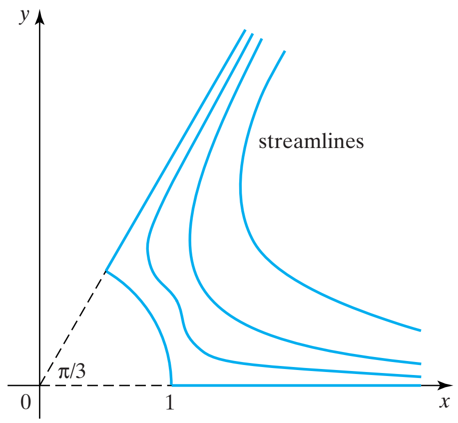


FIGURE 17.41 Flow in an indented wedge.

10. Find the complex potential for the flow inside the indented wedge shown in Fig. 17.42 when the flow moves parallel to each wall at infinity with speed U.

11. The Joukowski transformation $w = z + 1/z$ maps the upper half of the z-plane from which has been deleted a unit semicircle centered on the origin onto the upper half of the w-plane with a cut along the real axis from $w = -2$ to $w = 2$, as shown in Fig. 17.43. If w is the

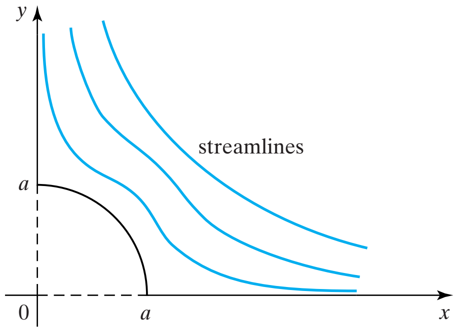


FIGURE 17.42 Flow in an indented right-angled wedge.

complex potential of a fluid flow, by setting $z = x + iy$ and $w = u + iv$, find the implicit equation of the streamlines in the z-plane corresponding to the flow lines $v = c (c \geq 0)$ in the w-plane. By examining the qualitative properties of the implicit equation of the streamlines, confirm they have the properties shown in Fig. 17.43, which can be interpreted as the flow of very deep water over a semicircular obstacle resting on the bottom. State how the diagram on the left can be used to describe the flow of a stream of water of finite depth over a submerged obstacle, when the surface of the stream is a *free surface* (a fluid–air interface).

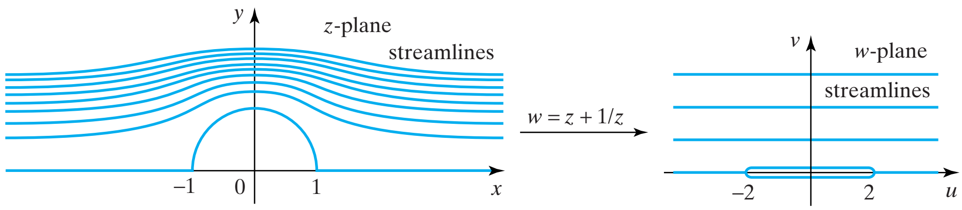


FIGURE 17.43 Flow over a semicircular obstacle.

12. The transformation $w = z + \exp z$ maps the strip $-\pi \leq y \leq \pi$ in the z-plane onto the w-plane with cuts along the lines $u \leq -1, v = \pm\pi$, as shown in Fig. 17.44. If w is the complex potential of a fluid flow, by setting $z = x + iy$ and $w = u + iv$ find the equation of the streamline $y = c$ in parametric form. As the cuts are bounded by streamlines, and fluid cannot cross a streamline, the cuts can be interpreted as parallel barriers, allowing the diagram on the right to be interpreted as flow emerging from a parallel channel into an unrestricted region. How can this problem be interpreted in terms of an electrostatic potential?

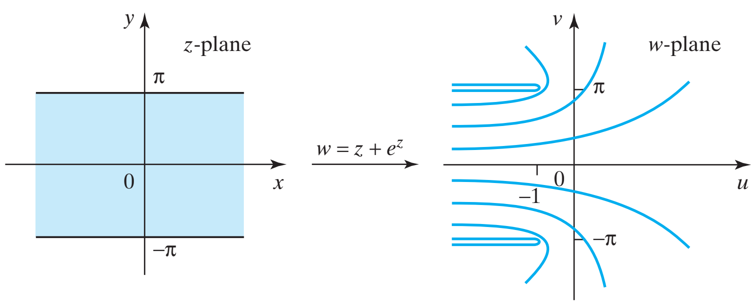


FIGURE 17.44 Flow from a parallel channel into an unrestricted region.

13. The transformation $w = \text{Arcsin}\, z$ maps the upper half of the z-plane with a cut along the real axis from $z = -1$ to $z = 1$, onto the semi-infinite strip $-\pi \leq u \leq \pi, v \geq 0$ in the

w-plane, as shown in Fig. 17.45. Use this result to find the equipotentials and flux lines if $A_\infty B$ is an electrically conducting plate at the constant electric potential $\phi = 200$, CD_∞ is an electrically conducting plate at the constant potential $\phi = 100$, and BC is an insulator (no flux lines can cross it). How can this problem be interpreted in terms of steady state heat conduction?

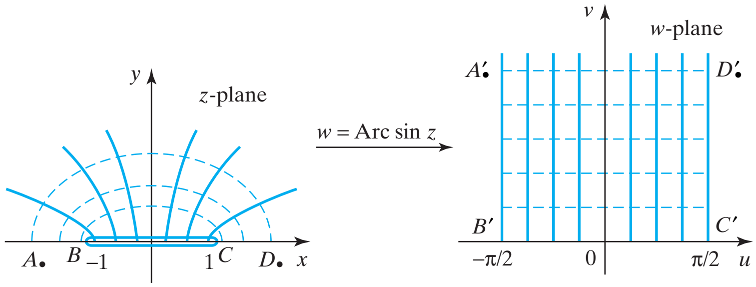


FIGURE 17.45 Electrically conducting plates separated by an insulator.

14. The diagram in Fig. 17.46 represents a metal lamina occupying the first quadrant of the (x, y)-plane with the edge $x = 0$, $y > 1$ maintained at the constant temperature $T = 200$, the edge $x > 1$, $y = 0$ maintained at the constant temperature $T = 50$, and the edges $x = 0$, $0 < y < 1$ and $0 < x < 1$, $y = 0$, maintained at the constant temperature $T = 0$. Find the temperature $T(x, y)$ at any point (x, y) in the lamina.

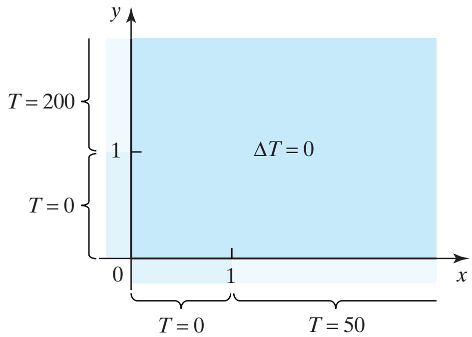


FIGURE 17.46 Mapping the exterior of a hole onto a half-plane.

15. The diagram on the left of Fig. 17.47 shows a cross-section of an infinite metal block pierced by a hole of unit diameter, the boundary DAB of which is maintained at the constant temperature $T = 450$, while the boundary DCB is maintained at the constant temperature $T = 100$. Use the fact that the transformation

$$w = i\left(\frac{z+1}{(1-i)z-1-i}\right)$$

maps $|z| \geq 1$ onto the upper half of the w-plane in the manner shown in Fig. 17.47 to find the temperature and isothermals in the plate.

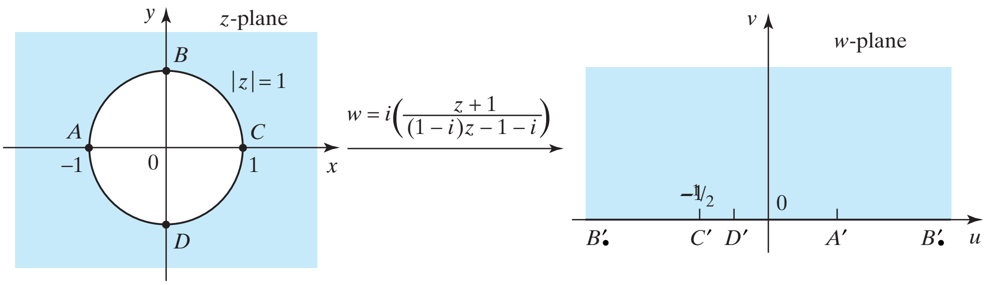
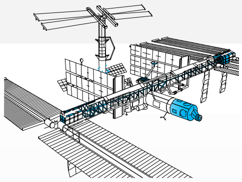


FIGURE 17.47 A metal block pierced by a hole.



CHAPTER 17
TECHNOLOGY PROJECTS

Project 1

Examining the Mapping of Lines and Circles by the Linear Fractional Transformation

The purpose of this project is to apply computer algebra and graphics to the linear fractional transformation

$$w = \frac{2z - i}{z + i}$$

to explore how it maps various straight lines and circles in the z-plane onto lines and circles in the w-plane, though not necessarily in this order.

Find the maps of (a) $y = 0$, (b) $y = 2$, (c) $y = x$, (d) the circle of radius $\frac{1}{2}$ centered on the origin, (e) the unit circle centered on the origin, and (f) the unit circle centered on the point $z = 1 + i$.

Project 2

This project examines the way the Joukowski transformation

$$w = z + \frac{1}{z}$$

maps a circle of radius R passing through the point $z = -1$ with its center in the first quadrant.

Experiment by choosing different positions for the center of the circle and then mapping the boundary of the circle onto the w-plane.

Project 3

Verify the results of Example 17.4 by using the function

$$w = \frac{z - 2 + \sqrt{3}}{(2 - \sqrt{3})z - 1}$$

to plot the map of the circles $|z - \frac{1}{4}| = \frac{1}{4}$ and $|z| = 1$ in the z-plane onto the w-plane. Hence, show that the circles map onto the concentric circles shown in Fig. 17.34.

Project 4

By considering the way $w = z + \exp(z)$ maps the infinite strip $-\pi \leq y \leq \pi$ in the z-plane onto the w-plane, show how this mapping can be interpreted as the two-dimensional discharge of fluid from between parallel semi-infinite planes into a surrounding infinite volume of fluid. Find the slope of the fluid flow lines far from the place of discharge, and plot some representative flow lines.

Explain how this same mapping can describe equipotentials inside and outside a parallel plate capacitor in a vacuum when the lower plate is at a potential V_1 and the upper plate is at a potential V_2, and determine the potential associated with each equipotential.

Project 5

In two-dimensional fluid mechanics, a **line source** of strength m is a line normal to the plane of the flow from which fluid enters the surrounding medium symmetrically at a steady rate of m volume units per unit line length per unit time. Similarly, when m is negative, this becomes a **line sink** that removes fluid from the surrounding medium symmetrically at a steady rate of m volume units per unit line length per unit time.

By considering the fluid complex potential

$$w = \phi + i\psi = m\,\mathrm{Log}(z - z_0), \quad (m > 0)$$

find the curves $\phi =$ constant and $\psi =$ constant that are, respectively, the equipotentials and streamlines of the flow. Hence, explain why w is the fluid complex potential of a line source located at a point z_0, with the line source perpendicular to the z-plane.

If attention is confined to the upper half of the z-plane, explain why the function

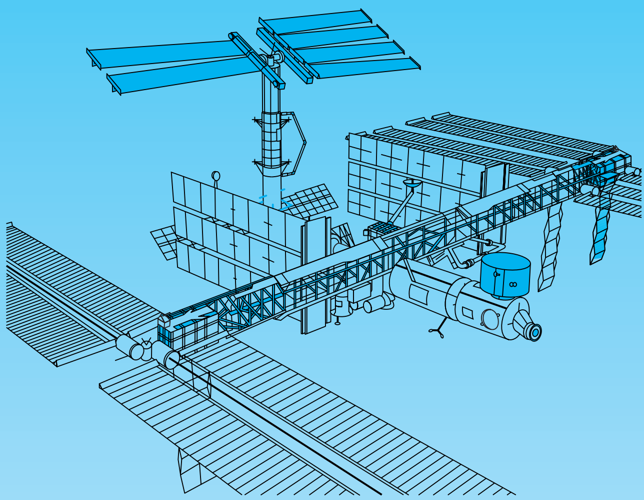
$$w = m \operatorname{Log}\left(\frac{z - z_0}{z - \bar{z}_0}\right)$$

is the complex potential for fluid flow in the upper half of the z-plane due to a line source of strength m located at z_0 when the region is bounded below by a fixed impenetrable barrier along the x-axis.

Plot the equipotentials and streamlines for such a flow for $-3 \leq x \leq 3, 0 \leq y \leq 3$ when $m = 1$, $z_0 = i$.

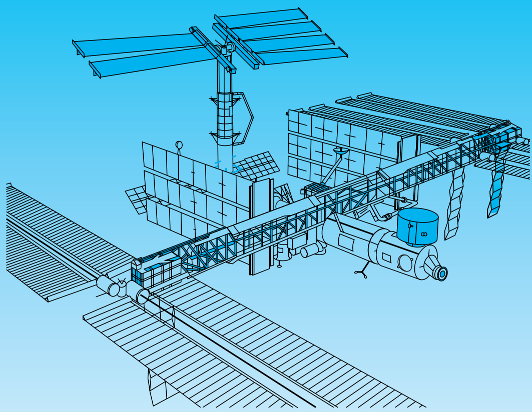
PART SEVEN

PARTIAL DIFFERENTIAL EQUATIONS



This Page Intentionally Left Blank

CHAPTER 18



Partial Differential Equations

Partial differential equations (PDEs) are equations satisfied by partial derivatives of functions of two or more independent variables. They describe all types of physical phenomena in engineering and science, ranging from transient heat conduction through vibrations of strings and plates to fluid flow and the behavior of electric and magnetic fields. The solution of first order equations is developed using the method of characteristics, and the three fundamentally different types of second order PDE are derived from first principles using typical physical examples. After classifying second order equations and describing suitable boundary and initial conditions, it is shown how the PDEs can be reduced to their standard forms to simplify the task of finding a solution. The wave equation is interpreted in terms of two disturbances propagating with equal speed, but in opposite directions, and the D'Alembert solution is derived.

The separation of variables method of solution is developed and related to the Sturm–Liouville systems, eigenvalues, and eigenfunctions already discussed in connection with ordinary differential equations. The method is then applied to various physical problems involving cartesian, cylindrical, and spherical polar coordinates. Some results of general importance to the study of PDEs are derived, and the chapter ends with an introduction to Laplace and Fourier transform methods of solution for PDEs.

18.1 What Is a Partial Differential Equation?

The simplest form of partial differential equation (PDE) involving a suitably differentiable unknown function (**dependent variable**) $u(x, y)$ of the two independent variables x and y is an equation that relates x, y, u, and some partial derivatives of u with respect to x and y. The **order** of the PDE is the order of the highest partial derivative of u that occurs in the equation, so a general *first order* PDE for the function $u(x, y)$ is of the form

order of a PDE

$$F(x, y, u, u_x, u_y) = 0, \tag{1}$$

where F is an arbitrary function of its arguments.

More generally, a first order PDE for a function $u(x_1, \dots, x_n)$ of the n independent variables $x_1, \dots, x_n$ is an equation of the form

$$G(x_1, \dots, x_n, u, u_{x_1}, \dots, u_{x_n}) = 0, \tag{2}$$

where G is an arbitrary function of its arguments and $u_{x_i} = \partial u/\partial x_i$, for $i = 1, 2, \dots, n$.

First order equations are of special interest because they occur frequently in practical problems. Furthermore, from among all possible classes of PDE, they are the ones that are simple enough to permit study in great detail, and for which methods of solution exist that extend to certain types of second order equation.

A general *second order* PDE for a function $u(x, y)$ of the two independent variables x and y is of the form

$$H(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0, \tag{3}$$

where H is an arbitrary function of its arguments, and for conciseness the suffix notation $u_x = \partial u/\partial x, u_y = \partial u/\partial y, u_{xx} = \partial^2 u/\partial x^2, u_{yx} = \partial^2 u/\partial x\partial y$, and $u_{yy} = \partial^2 u/\partial y^2$ has been used.

classical and generalized solutions

A **classical solution** of a PDE defined in some region D of the (x, y)-plane is a *real* function u with the property that all of its partial derivatives that occur in the PDE are defined and *continuous* throughout D, and when the function is substituted into the PDE it satisfies the equation identically. We will see later that in certain cases a slightly more general class of solution is also possible where a derivative may be discontinuous. Solutions of this type are called **generalized solutions**, and they are often used in connection with wave propagation problems.

The expressions in (1) and (2) are too general to be directly useful, so only some important special cases will be examined. In the case of (1) the three special cases to be considered are called, respectively, first order PDEs of *linear*, *semilinear*, and *quasilinear* type.

The Linear First Order PDE for $u(x, y)$

A *linear first order* PDE for the unknown function $u(x, y)$ can always be written as

$$p(x, y)u_x + q(x, y)u_y = r(x, y)u + s(x, y), \tag{4}$$

where $p(x, y)$, $q(x, y)$, $r(x, y)$, and $s(x, y)$ are arbitrary functions of x and y, and the term $s(x, y)$ that does not multiply u, u_x, or u_y is called the **nonhomogeneous** term. The PDE is called **homogeneous** when $s(x, y) = 0$. When, as often happens, the functions p, q, and r are constants, the PDE becomes a **constant coefficient** equation. The equation in (4) is called *linear* because u, u_x, and u_y all occur linearly (with degree 1) in each term. The following is a typical linear first order PDE:

$$u_x + xu_y = u + 2.$$

The solution $u = u(x, y)$ of (4) in a region D of the (x, y)-plane where the PDE is defined can be represented in the form of a surface above D called an **integral surface**. For most PDEs it is impossible to find a general solution so instead, when solving a PDE, it is usual to consider a specific problem by requiring that as

well as the solution satisfying the PDE, it also satisfies some auxiliary (additional) conditions that identify the particular problem.

In the case of a linear first order PDE it will be seen later that in principle a general solution can be found, though usually only the solution of a specific problem is required. In order to specify such a problem for a first order PDE, the auxiliary condition that identifies the problem uniquely involves prescribing the value the solution u is required to attain along a line in D. An auxiliary condition of this nature is called a **Cauchy condition**, and the problem of finding the solution of a PDE in D that satisfies a Cauchy condition is called a **Cauchy problem** for the PDE. More will be said about the Cauchy problems in the next section.

Cauchy conditions

The Semilinear First Order PDE for $u(x, y)$

A *semilinear first order* PDE is slightly more complicated than a linear first order equation because it is of the form

$$p(x, y)u_x + q(x, y)u_y = f(x, y, u), \tag{5}$$

where f is an arbitrary *nonlinear* function of u. The left sides of the PDEs in (4) and (5) are identical, but the right side of the semilinear PDE in (5) depends *nonlinearly* on u instead of linearly as in (4). A typical example of a semilinear first order PDE is

linear, semilinear, and quasilinear first order PDEs

$$u_x + (1 + x)u_y = (1 + x + y)u^2,$$

where the term $f(x, y, u) = (1 + x + y)u^2$ is nonlinear because of the term u^2.

The Quasilinear First Order PDE

A *quasilinear first order* PDE is one that can be written in the form

$$p(x, y, u)u_x + q(x, y, u)u_y = f(x, y, u) \tag{6}$$

where the functions p and q may or may not depend on x and y, but at least one of them depends on the undifferentiated function u. When f is present in (6) it may or may not depend on all of x, y, and u, though the presence or absence of f does not alter the quasilinear nature of the equation. A typical quasilinear first order PDE is

$$u_x + uu_y = u,$$

where in this case the quasilinearity is due to the presence of the term uu_y.

Both linear and quasilinear first order PDEs often occur in systems involving several dependent variables, and on occasion it is possible for all but one of the dependent variables to be eliminated, leading to a single higher order equation in the remaining dependent variable. The following is an example of a simple linear system of first order equations involving the variables $v(x, t)$ and $w(x, t)$:

$$v_t - c^2 w_x = 0 \quad \text{and} \quad w_t - v_x = 0. \tag{7}$$

Here c is a constant. In these equations the independent variables are denoted by x and t, because in physical problems governed by these equations x is usually a space variable (a length) and t is the time.

When v and w are twice differentiable functions, partial differentiation of the first equation with respect to t gives

$$v_{tt} - c^2 w_{xt} = 0,$$

and partial differentiation of the second equation with respect to x gives

$$w_{tx} - v_{xx} = 0.$$

Provided the second derivatives are continuous, the mixed derivatives are equal, so that $w_{xt} = w_{tx}$. After the elimination of w_{xt} between these two equations, the following linear second order equation for v is obtained:

$$v_{tt} - c^2 v_{xx} = 0. \tag{8}$$

Had the first equation in (7) been differentiated partially with respect to x and the second equation partially with respect to t, this same argument would have given

$$w_{tt} - c^2 w_{xx} = 0,$$

showing that v and w both satisfy the same PDE.

Later this equation will be seen to describe an important form of wave propagation in one space dimension and time, and for this reason it is called the *one-dimensional wave equation*. In the wave equation the constant c is the speed with which waves (disturbances) are propagated. Another linear example is provided by the *Cauchy–Riemann equations* (see Section 13.2)

$$u_x = v_y \quad \text{and} \quad u_y = -v_x,$$

where u and v are the real and imaginary parts of an analytic function $f(z) = u + iv$, with $z = x + iy$. In this case an argument similar to the one just used shows that both u and v are harmonic functions, so as each is a solution of Laplace's equation,

$$u_{xx} + u_{yy} = 0 \quad \text{and} \quad v_{xx} + v_{yy} = 0.$$

A more complicated system of quasilinear equations is provided by the equations of *unsteady* (time dependent) *gas dynamics*. In their simplest form these equations relate the gas density ρ, its pressure $p = k\rho^\gamma$ with k and γ constants, and the gas velocity $\mathbf{u}$, all at time t and at some position vector $\mathbf{r}$ in space, through the system of equations

$$\rho_t + \operatorname{div}(\rho\mathbf{u}) = 0 \quad \text{and} \quad \mathbf{u}_t + \mathbf{u}\cdot\nabla\mathbf{u} + (1/\rho)\nabla p = 0. \tag{9}$$

The first equation is a scalar equation that describes the conservation of mass, and the second is a vector equation with three scalar components that is related to the equation that describes the conservation of momentum. The system in (9) couples the density ρ and the three scalar components of $\mathbf{u}$ through a system of four scalar quasilinear equations. In this case the structure of the system is such that it cannot be replaced by a single higher order equation for one of the unknowns.

When introducing the linear first order PDE, mention was made of the fact that the complexity of PDEs is such that general solutions can only be found in very special cases. As a result, when dealing with higher order PDEs, instead of seeking general solutions, methods are developed that enable solutions of specific problems to be found. As already mentioned, to find the solution of a particular problem involving a PDE it is necessary to require that the solution satisfy some auxiliary

conditions that identify the problem. The additional conditions may be imposed on spatial boundaries belonging to a region D where the solution is required, and when this is done the conditions are called **boundary conditions**. A typical boundary condition for a second order PDE defined in a rectangle could be that the solution is required to assume specified values on the sides of the rectangle. If time is involved, it is necessary to specify how the solution starts, and a condition of this type is called an **initial condition**. Problems requiring initial and boundary conditions are called **initial boundary value problems** (IBVPs).

boundary and initial conditions

The definitions of linearity and quasilinearity extend quite naturally to PDEs of all orders. A PDE of any order is **linear** if the unknown function u and all its derivatives only appear linearly (to degree 1), so a general linear second order PDE for the unknown function $u(x, y)$ can be written

$$a(x, y)u_{xx} + b(x, y)u_{xy} + c(x, y)u_{yy} + d(x, y)u_x + e(x, y)u_y + f(x, y)u = h(x, y). \tag{10}$$

Analogously, a PDE of order n is said to be **quasilinear** when its partial derivatives of order n occur linearly in the equation, but combinations of u and some of its derivatives up to order $n-1$ occur as coefficients of the nth order partial derivatives. A general quasilinear second order PDE for the unknown function $u(x, y)$ can be written

linear, quasilinear, and nonlinear higher order PDEs

$$\begin{aligned} &a(x, y, u, u_x, u_y)u_{xx} + b(x, y, u, u_x, u_y)u_{xy} \\ &\quad + c(x, y, u, u_x, u_y)u_{yy} + h(x, y, u, u_x, u_y) = 0, \end{aligned} \tag{11}$$

where a, b, c, and h are arbitrary functions of their arguments, with at least one of the functions a, b, and c depending on u and/or one or more of its first order partial derivatives.

A PDE of any order that is not linear, semilinear, or quasilinear is said to be **nonlinear**. The following is an example of a nonlinear second order PDE:

$$uu_{xx} + \sin(u_{yy}) + xu_x + u_y + u = 0.$$

Here the nonlinearity is caused by the term $\sin(u_{yy})$.

Although in principle a general solution of a linear first order PDE can be found, unlike the general solution of a linear first order ordinary differential equation (ODE) that contains an arbitrary *constant*, the general solution of a linear first order PDE contains an arbitrary *function*. This situation is illustrated by the first order PDE

$$u_x + xu_y = u + 2, \tag{12}$$

which can be shown to have the general solution

$$u(x, y) = C\exp\{x + \phi(\xi)\} - 2, \tag{13}$$

where $\xi^2 = x^2 - 2y$, ϕ is an arbitrary differentiable function of its argument ξ and C is a constant.

To find a specific solution suppose, for example, that a solution of (12) is required to satisfy the auxiliary condition $u(x, 0) = -1$. Setting $y = 0$ in the general solution, and noticing that as $\xi^2 = x^2 - 2y$ it follows that on the x-axis $\xi = x$, we find from the condition $u(x, 0) = -1$ that the arbitrary function ϕ must be chosen such that

$$-1 = C\exp\{x + \phi(x)\} - 2, \qquad \text{and so } 1 = C\exp\{x + \phi(x)\}.$$

This is only possible if $C = 1$ and $\phi(x) = -x$, so replacing x in $\phi(x)$ by $\xi = (x^2 - 2y)^{1/2}$ gives $\phi(\xi) = -(x^2 - 2y)^{1/2}$, so the solution becomes

$$u(x, y) = C \exp\{x - (x^2 - 2y)^{1/2}\} - 2.$$

Differentiation confirms that this expression satisfies the PDE, so as it also satisfies the additional condition $u(x, 0) = -1$ it is the required classical solution. The solution will be real provided $x^2 \geq 2y$, so the line $y = 0$ on which the Cauchy condition is specified is seen to bound the region of the (x, y)-plane where the classical solution is defined.

existence and uniqueness

Two important questions that must be answered when working with PDEs are (i) the **existence** question (does the PDE have a solution?) and (ii) the **uniqueness** question (if a solution exists, is it the only possible one?).

These questions can be answered in some detail for first order PDEs and higher order linear equations, and to a lesser extent for other types of PDEs, but it will suffice to say here that a solution of a linear PDE exists, and when the additional condition in the form of a Cauchy condition is specified in a manner to be described later, the corresponding solution will be unique.

To see that not every first order PDE has a solution, it is only necessary to consider the nonlinear equation

$$u_x^2 + u_y^2 = -1.$$

The expression on the left is nonnegative, so clearly this equation cannot be satisfied by any *real* function $u(x, y)$.

derivation of the first order PDE involving a transient heat balance

To illustrate one of the ways in which first order PDEs arise from physical situations, we will derive the equation governing the transient heat balance between a pipe transporting a hot fluid and the air surrounding the pipe at a constant temperature T_0. Let the length of the pipe be L, the constant speed of the fluid through the pipe be u, and the temperature of the fluid be $T(x, t)$, where x is the distance along the pipe and t is the time measured from the moment a particle of fluid enters the pipe. The physical situation is represented in Fig. 18.1, and in order to arrive at the transient heat balance equation we will consider the situation in an element of the pipe of length Δx.

The instantaneous energy balance that is to be modeled in the element of pipe of length Δx can be represented as follows:

{energy entering with fluid}$-${energy leaving with fluid}$-${heat transferred to air}
$=${energy stored in fluid}.

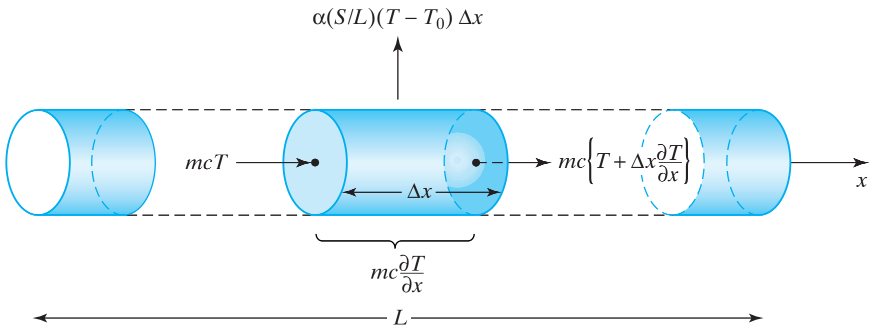


FIGURE 18.1 Transient heat distribution in an element of the pipe of length Δx.

If Δt is the time taken for a particle of fluid to travel through an element of the pipe of length Δx, the fluid speed $u \approx \Delta x/\Delta t$. If we denote the mass of fluid present in this element by M and the mass flow rate by m, the quantities M and m are related by $M = m\Delta x/u$.

If the fluid enters the element at the temperature $T(x, t)$, its temperature when leaving it can be approximated by $T + \Delta x(\partial T/\partial x)$. If we assume that the transfer of heat from the surface of the pipe to the air is proportional to the temperature difference $T(x, t) - T_0$, and denote the heat transfer coefficient by α, the heat transferred from the surface of the pipe to the air will be $(\alpha S\Delta x/L)(T - T_0)$, where S is the surface area of the pipe. The heat energy entering the element due to the fluid is mcT, where c is the specific heat of the fluid, and the heat energy leaving with the fluid is $mc(T + \Delta x\partial T/\partial x)$, whereas the stored energy in the fluid occupying the element is $Mc(\partial T/\partial t)$. Substituting these quantities into the energy balance equation gives

$$mcT - mc\left(T + \Delta x\frac{\partial T}{\partial x}\right) - \alpha\left(\frac{S\Delta x}{L}\right)(T - T_0) = Mc\frac{\partial T}{\partial t}. \tag{14}$$

Cancelling terms, and dividing (14) by $Mc = cm\Delta x/u$, this balance equation becomes the PDE for transient heat transfer:

$$\boxed{\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} = -\frac{\alpha u S}{mcL}(T - T_0).} \tag{15}$$

Other examples of the derivation of PDEs that govern the behavior of important but very different physical situations are to be found in Section 18.5 where the three fundamental types of linear second order PDE are derived.

Summary

First and second order partial differential equations (PDEs) of linear, quasilinear, and nonlinear type have been defined. The Cauchy problem has been introduced and the questions of the existence and uniqueness of solutions raised. A typical first order PDE has been derived from a physical problem involving the transient heat balance between a pipe carrying hot water and the surrounding air.

EXERCISES 18.1

Classify the PDEs in Exercises 1 and 2 as linear, semilinear, quasilinear, or nonlinear.

1. (a) $u_x + u^2u_y = x + 2y$.
(b) $3u_x + 4u_y = \sin x$.
(c) $u_x + xu_y^2 = u + 1$.
(d) $u_x + 2u_y = \cos u$.
(e) $(x + 1)u_x + yu_y = 2u + e^x$.
(f) $u_x + (1 + u_x)u_y = u^2$.
(g) $(x^2 + 1)u_{xx} - yu_{yy} = 1 + \cos x$.
(h) $u_{xx} + (1 + u_x^{3/2})u_{yy} = \sin u$.

2. (a) $u_x \sin y + u_y \cos x = 1 + x^2 + y$.
(b) $u_x + (1 + u)u_y = 2xy$.
(c) $(x^2 + 1)u_x + u_y^2 = 2x + 3$.
(d) $(1 + x + x^2)u_x + (2y + 1)u_y = 1$.
(e) $(xy + 2)u_x + (1 + y + u)u_y = u$.
(f) $u_x \sin x + u_y \cos y = x + y + 3u$.
(g) $u_{xx} - u_{yy} = \sin u$.
(h) $u_{xx} - 2xu_{xy} + (1 + \cos u)u_{yy} = 4$.

In Exercises 3 through 6 use the general solution of the PDE in (12) given in (13) to find the solution that satisfies the given condition, stating any restriction that is required for the solution to be valid.

3. $u(x, 0) = 2,\ y > 0$.
4. $u(x, 0) = e^{2x} - 2,\ y > 0$.
5. $u(x, 1) = -1,\ y > 0$.
6. $u(x, 2) = x - 2,\ y > 2$.

18.2 The Method of Characteristics

The method of solution of a quasilinear first order PDE involving the unknown function $u(x, y)$ contains within it as special cases the solution of linear and semilinear first order PDEs. Consequently it is only necessary to discuss the solution of a Cauchy problem for a quasilinear equation that we will write in the form

$$p(x, y, u)u_x + q(x, y, u)u_y = f(x, y, u), \tag{16}$$

where p, q, and f are assumed to be continuous functions of their arguments. The Cauchy condition for u will be imposed on a curve Γ in the (x, y)-plane on which u will be required to assume a prescribed functional form, with the function depending on the position on Γ.

Cauchy data curve, initial line

When the independent variables x and y are space variables, the curve Γ will be called the **Cauchy data curve**. If, however, one independent variable is a space variable and the other is the time, and Γ coincides with the x-axis, it is natural to refer to Γ as the **initial line** and to the Cauchy condition itself as the **initial condition** (or the **initial data**) for the PDE. It is then understood that as time increases the solution will evolve away from the initial condition.

If the Cauchy data curve Γ is complicated, it is usually necessary to define it parametrically by writing

$$x = x_0(s), \quad y = y_0(s), \tag{17}$$

for all values of a parameter s in some appropriate interval I. So, for example, if Γ is the straight line through the origin $ax - by = 0$, one possible parametrization of the line involves setting $x = bs$ and $y = as$ for $-\infty < s < \infty$.

In (17) the functions $x_0(s)$ and $y_0(s)$ are assumed to be continuous with piecewise continuous derivatives $x_0'(s)$ and $y_0'(s)$ such that $(x_0'(s))^2 + (y_0'(s))^2 \neq 0$. This last condition ensures that the length element $dl = \sqrt{\{(x_0'(s))^2 + (y_0'(s))^2\}}ds$ along Γ increases steadily with s. We will see later that the Cauchy data curve Γ cannot be specified in a completely arbitrary manner, and the nature of the restriction that must be placed on it will become clear when the method of solution has been developed.

When Γ has been defined parametrically in terms of s, the initial condition $u = u_\Gamma$ on Γ can also be defined in terms of s by setting

$$u_\Gamma(s) = u_0(s), \tag{18}$$

where $u_0(s) = u_0(x_0(s), y_0(s))$ is a prescribed function.

The total derivative of a function $u(x, y)$ along an arbitrary curve defined parametrically in terms of a parametric variable σ by the differentiable functions $x = x(\sigma)$, $y = y(\sigma)$ is

$$\frac{du}{d\sigma} = \frac{\partial u}{\partial x}\frac{dx}{d\sigma} + \frac{\partial u}{\partial y}\frac{dy}{d\sigma}. \tag{19}$$

A comparison of (16) and (19) shows that by setting

$$\frac{dx}{d\sigma} = p(x, y, u) \quad \text{and} \quad \frac{dy}{d\sigma} = q(x, y, u), \tag{20}$$

the PDE in (16) can be expressed as the ODE

$$\frac{du}{d\sigma} = f(x, y, u), \tag{21}$$

provided x and y satisfy (20).

characteristic equations, characteristics, and the compatibility condition

The two ODEs in (20) are called the parametric form of the **characteristic equations** of the PDE in (16), and when they are integrated to obtain an expression of the form

$$\Phi(x, y, k) = 0, \tag{22}$$

where k is a constant of integration, they define a family of curves C in the (x, y)-plane called the **characteristic curves** of the PDE, each of which is identified by a different value of k. Notice that in quasilinear PDEs the characteristics depend on the solution u, so in such cases it is necessary to solve (20) and (21) simultaneously. For conciseness, the curves belonging to the family C are usually called the **characteristics** of the PDE. The ODE in (21) is called the **compatibility condition** along the characteristic.

If required, the parameter σ can be eliminated from the characteristic equations and the compatibility condition by dividing the second ODE in (20), and the ODE in (21), by $dx/d\sigma$ given in the first of the equations in (20). This leads to the equation for the *characteristic curves*

$$\frac{dy}{dx} = \frac{q(x, y, u)}{p(x, y, u)} \tag{23}$$

and to the *compatibility condition*

$$\frac{du}{dx} = \frac{f(x, y, u)}{p(x, y, u)}. \tag{24}$$

Although the equations (23) and (24) appear simpler than the equivalent ones in (20) and (21), in many cases the equations in terms of the parameter σ are easier to integrate.

The representation of the PDE in (16) as the set of ODEs in (20) and (21) or, equivalently, as the ODEs in (23) and (24) forms the basis of a method of solution for a first order PDE for $u(x, y)$ called the **method of characteristics**.

method of characteristics

The significance of the characteristic curves and the compatibility condition is most easily understood by considering the intersection of a representative characteristic curve and the Cauchy data curve Γ. Consider the characteristic curve C^* in Fig. 18.2 that intersects Γ at a point P corresponding to $s = s^*$ in the parametrization of Γ. As P is the point $(x_0(s^*), y_0(s^*))$, in the (x, y)-plane, the Cauchy condition at P is $u = u_0(s^*)$. The solution $u(x, y)$ of the PDE will then be determined along the characteristic curve C^* by integration of the compatibility condition (21) subject

to the initial condition $u = u_0(s^*)$, with similar interpretations when (23) and (24) are used.

It can be seen from this argument that when the PDE in (16) is either linear or semilinear, the characteristic curves can be determined *independently* of the solution by integrating either (20) or (23), because in these two cases the solution u does not enter into the functions p and q. Consequently, in these two cases, solving the PDE in (16) reduces to the integration of the ODEs that determine the family of characteristic curves C, followed by the integration of the compatibility condition along the characteristic curves subject to an appropriate initial condition. Figure 18.2 illustrates the application of the method of characteristics to linear and semilinear PDEs written in the form

$$p(x, y)u_x + q(x, y)u_y = f(x, y, u), \tag{25}$$

where f depends linearly on u when (25) is linear, and nonlinearly on u when it is semilinear.

If the PDE is quasilinear, the solution u enters into the equations determining the characteristics, so when this occurs the integrations can only be performed analytically when the equations involved are simple. In general, when working with quasilinear first order PDEs, and also with linear and semilinear PDEs with complicated coefficients, the system of ODEs comprising the characteristic equations and the compatibility condition must be solved simultaneously using a numerical integration technique such as the Runge–Kutta method described in Chapter 19.

The **uniqueness** of the solution $u(x, y)$ in (25) follows directly from the way in which the method of characteristics produces the solution, and the fact that integration along a typical characteristic C^* of the compatibility condition (see Fig. 18.2) leads to a solution for $u(x, y)$ that depends *uniquely* on the initial condition $u = u_0(s^*)$ associated with the characteristic. The solution will cease to be unique if intersection of characteristics occurs at a point Q in the (x, y)-plane. This is because, in general, the value of u at Q determined by integration of the compatibility condition along each of the characteristics that meet there cannot be expected to be in agreement.

The restriction that must be placed on the initial curve Γ can be seen by considering Fig. 18.3. Provided Γ is nowhere tangent to a characteristic, as is the case for the characteristic C_P through point P, the solution along C_P will evolve according to

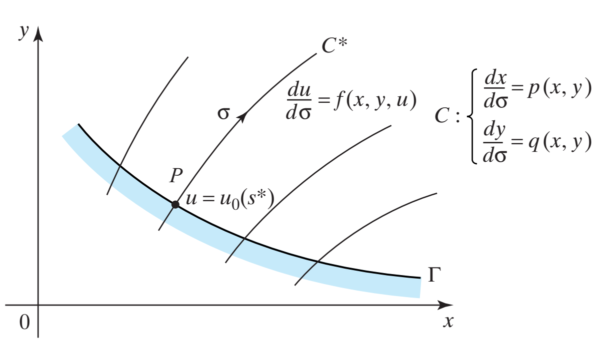


FIGURE 18.2 The solution of a linear or semilinear PDE by the method of characteristics.

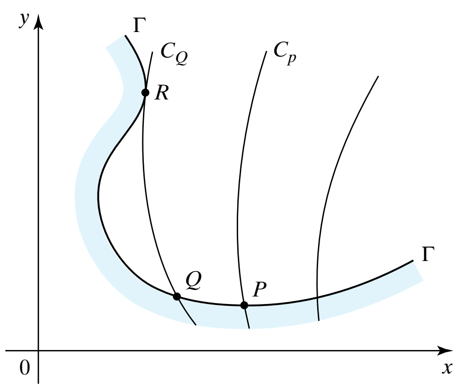


FIGURE 18.3 Tangency and nontangency of characteristic curves and the initial line Γ.

the solution of the compatibility condition subject to the initial condition $u = u_0(P)$. The situation is different, however, in the case of the characteristic curve C_Q through the point Q that becomes tangent to the Cauchy data curve Γ at point R.

In this case the Cauchy condition $u = u_0(R)$ specified at R where the Cauchy data curve Γ is tangent to C_Q cannot be expected to be in agreement with the solution obtained by integrating the compatibility condition along C_Q from Q to R subject to the initial condition $u = u_0(Q)$ at Q. This shows that when specifying a Cauchy problem for the PDE in (16) it is necessary that the initial curve Γ be nowhere tangent to a characteristic curve. As the characteristics can be determined independently of the solution u when the PDE is linear or semilinear, for such equations it is always possible to determine in advance that the nontangency condition is satisfied. If, however, the equation is quasilinear, then although the nontangency condition for Γ may be satisfied in neighborhood of Γ, this may not remain true as the solution evolves.

characteristic Cauchy problem

A special case of the Cauchy problem for the PDE in (16) arises when the Cauchy data curve Γ coincides with a characteristic curve of the equation. The determination of a solution for such a problem, when it exists, is called the **characteristic Cauchy problem**.

The following examples illustrate the application of the method of characteristics to linear, semilinear, and quasilinear first order PDEs, and also to a simple characteristic Cauchy problem. In general, equations (23) and (24) are the simplest to use when the Cauchy condition is prescribed on any straight line, and the parametric representation of the characteristic equations is only necessary when Cauchy data is prescribed on a curve. However, to illustrate the parametric approach, the second example makes use of equations (20) and (21) for the case where the Cauchy data is prescribed on a straight line through the origin.

Once a solution has been found it must always be checked to see that it satisfies both the prescribed Cauchy condition and the original PDE. The solution should also be examined to identify any restrictions that need to be placed on it in order to ensure that it remains real and finite.

EXAMPLE 18.1

Solve the Cauchy problem

$$u_x + 3u_y = 2u, \quad \text{given that } u(x, 0) = e^x.$$

Solution This is a linear equation, and as the Cauchy data curve is the x-axis, we will use the characteristic equations given in (23) and (24).

From (23) the characteristic curves of the PDE are determined by $dy/dx = 3$, so integration shows their equation to be $y = 3x + \xi$, where ξ is a constant of integration that corresponds to the point of intersection $(0, \xi)$ of the characteristic and the x-axis.

The compatibility condition is $du/dx = 2u$, so integration shows that

$$\ln u = 2x + f(\xi),$$

where $f(\xi)$ represents the arbitrary constant introduced as a result of the integration. This constant depends on the characteristic involved, but as a characteristic depends on ξ because of its point of intersection $(\xi, 0)$ with the x-axis, it is necessary to introduce the constant (on a particular characteristic) as $f(\xi)$, where f is an arbitrary function.

Substituting $\xi = y - 3x$ into the solution for u gives

$$u(x, y) = \exp\{2x + f(y - 3x)\}.$$

To find the form of the arbitrary function f, we now make use of the Cauchy condition, which in this case is $u(x, 0) = e^x$. Setting $y = 0$ in the expression for $u(x, y)$ and imposing the Cauchy condition gives

$$e^x = \exp\{2x + f(-3x)\},$$

and after taking logarithms this becomes

$$-x = f(-3x), \text{ that is equivalent to } f(x) = \frac{1}{3}x.$$

Replacing x by $y - 3x$ in $f(x)$, we have $f(y - 3x) = \frac{1}{3}y - x$, so substituting $f(y - 3x)$ into the expression for $u(x, y)$ gives

$$u(x, y) = \exp\left\{x + \frac{1}{3}y\right\}.$$

This function satisfies the Cauchy condition and differentiation confirms that it is a solution of the original PDE, so it is a classical solution of the equation. Inspection shows the solution to be valid throughout the entire (x, y)-plane. A solution such as this that is valid without restriction on its independent variables is called a **global** solution. ■

EXAMPLE 18.2

Cauchy problems for linear, semilinear, and quasilinear PDEs

Solve the Cauchy problem

$3u_x + 2u_y = x$, given that $u(x, y) = 1$ on the line Γ with the equation $ax = by$.

Solution This is a linear equation, with the Cauchy data curve Γ a straight line through the origin, so to illustrate the parametric approach we will use the characteristic equations given in (20) and (21).

We parametrize Γ by setting $x = bs$, $y = as$, where $-\infty < s < \infty$. The characteristic curves (lines in this case) are determined by (20), which when integrated become

$$x = 3\sigma + k_1, \quad y = 2\sigma + k_2.$$

When $\sigma = 0$ we know that x and y lie on Γ, but then $x = bs$ and $y = as$, so it follows that $k_1 = bs$, $k_2 = as$, showing that

$$x = 3\sigma + bs, \quad y = 2\sigma + as.$$

Solving these expressions for s and σ gives

$$s = \frac{3y - 2x}{3a - 2b}, \quad \sigma = \frac{ax - by}{3a - 2b}, \qquad \text{for } 3a \neq 2b.$$

The compatibility equation (21) becomes

$$\frac{du}{d\sigma} = x, \text{ but } x = 3\sigma + bs,$$

so after integration

$$u(s, \sigma) = \frac{3}{2}\sigma^2 + b\sigma s + f(s),$$

where $f(s)$ represents the usual arbitrary additive integration constant. As the characteristic depends on the parameter s, the integration constant $f(s)$ is shown as a function of s. The Cauchy condition $u(x, y) = 1$ is imposed on Γ, corresponding to $\sigma = 0$ in the preceding expression, so setting $\sigma = 0$ and replacing $u(s, 0)$ by 1 we find that $1 = f(s)$ for all s, and so in terms of s and σ the solution is seen to be given by

$$u(s, \sigma) = \frac{3}{2}\sigma^2 + b\sigma s + 1.$$

Replacing s and σ by their expressions in terms of x and y, we arrive at the explicit solution in terms of x and y

$$u(x, y) = \frac{3}{2}\left(\frac{ax - by}{3a - 2b}\right)^2 + b\left(\frac{ax - by}{3a - 2b}\right)\left(\frac{3y - 2x}{3a - 2b}\right) + 1, \quad \text{for } 3a \neq 2b.$$

This function satisfies the Cauchy condition on Γ, and differentiation confirms that it satisfies the original PDE, so it is a classical solution of the equation. Inspection shows the solution to be valid in the entire (x, y)-plane provided $3a \neq 2b$, so it is a global solution if this condition is satisfied.

When $3a = 2b$, the preceding solution fails because the Cauchy data line Γ with the equation $2x - 3y = 0$ coincides with the characteristic through the origin, causing the problem to become a *characteristic Cauchy problem*.

To examine the solution in this case, we must allow for the fact that although both Γ and the characteristic through the origin coincide, they are each parametrized differently. From the equations defining the characteristics we have $dx/d\sigma = 3$ on the Cauchy data line $x = bs$, so $dx/ds = b$.

The compatibility condition is

$$\frac{du}{d\sigma} = x, \text{ so in terms of } s \text{ this can be written } \frac{du}{d\sigma} = bs.$$

To express the derivative on the left of this last result in terms of s we use the chain rule

$$\frac{du}{ds} = \frac{du}{d\sigma}\frac{d\sigma}{ds} = \frac{d\sigma}{dx}\frac{dx}{ds}\frac{du}{d\sigma}, \quad \text{and so } \frac{du}{ds} = \frac{b}{3}\frac{du}{d\sigma}.$$

Combining this result with $du/d\sigma = bs$ gives

$$\frac{du}{ds} = \frac{b^2}{3}s, \text{ and after integration this becomes } u = \frac{b^2}{6}s^2 + c, (c = \text{constant}).$$

Substituting $x = bs$ into this result, we arrive at the solution

$$u(x, y) = \frac{1}{6}x^2 + c.$$

This expression for $u(x, y)$ is a degenerate solution of the original PDE along the characteristic through the origin that coincides with the Cauchy data line. However, this is *not* a solution of the characteristic Cauchy problem, because it does not satisfy the Cauchy condition $u(x, y) = 1$ along the line Γ.

This shows that this characteristic Cauchy problem with the stated Cauchy condition along Γ has *no* solution. A solution for the characteristic Cauchy problem could only exist if the Cauchy condition on Γ is changed to $u(x, y) = \frac{1}{6}x^2 + c$.

This solution is not the most general one, because the fact that Γ has the equation $3y - 2x = 0$ allows us to add to the preceding solution any arbitrary differentiable function $f(3y - 2x)$ that is a solution of the *homogeneous* form of the PDE $3u_x + 2u_y = 0$, since the result will still be a solution. This shows that the most general solution of this characteristic Cauchy problem is

$$u(x, y) = \frac{1}{6}x^2 + f(3y - 2x),$$

provided this expression also satisfies the Cauchy condition on Γ. In this result the constant c that appeared earlier has been absorbed into the arbitrary function f.

This example demonstrates the fact that, in general, the characteristic Cauchy problem has no solution, but when it does the solution is not unique, because it contains an arbitrary function. ■

EXAMPLE 18.3 Solve the Cauchy problem

$$u_x + u_y = e^u, \text{ given that } u(0, y) = y.$$

Solution This is a semilinear equation, but this time the Cauchy condition is specified on the y-axis so it will be simplest to use the nonparametric form of the characteristic equations.

The characteristics are determined by

$$\frac{dy}{dx} = 1, \text{ and integration gives } y = x + \xi,$$

where ξ is the point $(0, \xi)$ on the y-axis through which the characteristic passes. The compatibility condition is

$$\frac{du}{dx} = e^u, \text{ and after integration this becomes } -e^{-u} + f(\xi) = x.$$

Here f, an arbitrary function of its argument ξ that identifies the characteristic as the one passing through the point $(0, \xi)$, again represents the arbitrary constant that enters as a result of the integration. Substituting $\xi = y - x$ into this last result gives

$$-e^{-u} + f(y - x) = x, \quad \text{or } u(x, y) = 1/\ln\{f(y - x) - x\}.$$

To find f we must now make use of the Cauchy condition $u(0, y) = y$. Setting $x = 0$, and replacing u by y, the preceding expression becomes

$$-e^y + f(y) = 0, \quad \text{so } f(y) = e^y,$$

from which it follows that $f(y - x) = e^{x-y}$. Substituting for $f(y - x)$ in the expression for $u(x, y)$, we find that

$$u(x, y) = \ln\left(\frac{1}{e^{x-y} - x}\right) \quad \text{for } e^{x-y} > x.$$

This expression satisfies the Cauchy condition specified, and differentiation confirms that it is a solution of the original PDE, so it is a classical solution. The restriction $e^{x-y} > x$ that ensures $u(x, y)$ is real shows that the solution is not defined over all of the (x, y)-plane, and so it is not a global solution. ■

EXAMPLE 18.4 Solve the Cauchy problem

$$u_x + uu_y + u = 0, \text{ given that } u(0, y) = 1 + y.$$

Solution This equation is quasilinear because of the presence of the term uu_y, and again the Cauchy condition is specified on an axis so the nonparametric form of the characteristic equations will be used.

The characteristic curves follow by integration of the equation

$$\frac{dy}{dx} = u,$$

on which the compatibility condition that determines u is

$$\frac{du}{dx} = -u.$$

Let the solution along the characteristic through the point $(0, \xi)$ on the y-axis be $u = g(\xi)$. Then integration of the compatibility condition along the characteristic with respect to x gives

$$\ln u = -x + \ln g(\xi), \quad \text{so} \quad u = g(\xi)e^{-x}.$$

It follows from the Cauchy condition that $u = 1 + \xi$ at the point $(0, \xi)$, so setting $x = 0$ and replacing u by $1 + \xi$ in this last result, we find that

$$g(\xi) = 1 + \xi,$$

and so

$$u = (1 + \xi)e^{-x}.$$

The equation determining the characteristic curves now follows if we use this last result in the equation $dy/dx = u$, to obtain

$$\frac{dy}{dx} = (1 + \xi)e^{-x}.$$

Integration of this result using the fact that the characteristic passes through the point $(0, \xi)$ leads to the result

$$y = \xi + (1 + \xi)\int_0^x e^{-\eta}d\eta,$$

so

$$y = (1 - e^{-x}) + \xi(2 - e^{-x}).$$

When ξ is eliminated the solution becomes

$$u = \left(\frac{1 + y}{2 - e^{-x}}\right)e^{-x}, \text{ provided } x \neq -\ln 2.$$

This function satisfies the Cauchy condition, and differentiation shows that it satisfies the original PDE, so it is a classical solution. The solution is not defined everywhere because it becomes infinite when $x = -\ln 2$. ■

Summary

This section introduced the method of characteristics for first order PDEs involving a scalar function of two independent variables. The method was seen to involve replacing the single PDE by two coupled ordinary differential equations (ODEs), one of which determined

the family of characteristic curves, while the other determined the variation of the solution along the characteristic curves. The method was seen to apply to linear, semilinear, and quasilinear equations, and in the linear and semilinear cases the characteristic curves could be determined independently of the solution. However, in the quasilinear case, the equations for the characteristics and for the variation of the solution along the characteristics had to be solved simultaneously.

EXERCISES 18.2

In Exercises 1 through 18 solve the given Cauchy problem. Verify that the result obtained is a solution, and comment on any restrictions that need to be placed on it.

1. $u_x + 2u_y = 2,\ u(x, 0) = x.$
2. $3u_x + 2u_y = x,\ u(x, 0) = 1.$
3. $4u_x + 3u_y = 1,\ u(x, y) = 3$ on $y = x.$
4. $yu_x + 3u_y = y(1 + u),\ u(0, y) = y^2.$
5. $2u_x + u_y = \cos x,\ u(x, 0) = \dfrac{3}{2}\sin x.$
6. $u_x + u_y = u - 1,\ u(x, 0) = 2x.$
7. $u_x + 2u_y = 2x,\ u(x, y) = 2$ on $y = 3x + 1.$
8. $u_x + xu_u = u + 2,\ u(0, y) = 3y.$
9. $u_x + 4xu_y = 3 + 2x\sec^2 x^2,\ u(x, 0) = 3x.$
10. $yu_x + u_y = y(u + 4),\ u(x, 0) = e^{2x}.$
11. $u_x + u_y = u^2,\ u(0, y) = y.$
12. $u_x + 2u_y = (1 + 2x)e^{-u},\ u(x, y) = 1$ on $y = x.$
13. $u_x + uu_y + u = 0,\ u(0, y) = \sin y.$
14. Obtain the solution in Example 18.2 without parameterizing the line $ax = by$.
15. $u_x + 2uu_y + 3u = 0,\ u(0, y) = 4y.$
16. $u_x + uu_y + u = 0,\ u(0, y) = e^y.$
17. $u_x + uu_y + u = 0,\ u(0, y) = 3 + 2y.$
18. $u_x + 2uu_y + 3u = 0,\ u(0, y) = 4y.$

18.3 Wave Propagation and First Order PDEs

A first order PDE for the unknown function $u(x, t)$ of two independent variables x and t of the form

$$u_t + p(x, t, u)u_x + q(x, t, u) = 0, \tag{26}$$

wave propagation and hyperbolic PDEs

where x has the dimensions of length and t is the time, can be considered to describe **wave propagation**. Here the term *wave* is used to describe an identifiable disturbance such as a sound or water wave that propagates at a finite speed through space as time increases. The PDE in (26) is called a first order **hyperbolic equation** because, like the second order wave equation to be considered later, it describes wave propagation. A typical equation of this type characterizing a physical problem was derived in Section 18.1, where a linear first order PDE was shown to model the transient heat flow from a pipe transporting a hot fluid.

To understand the different types of wave propagation that can be described by hyperbolic equations such as (26), it will be necessary to examine some typical cases. The method of solution that will be used is the method of characteristics described in Section 18.2. However, this time the variable x will be replaced by t, as it represents the time, and the variable y will be replaced by x, which represents a length. A Cauchy condition for (26) specified at some fixed time, typically $t = 0$, is an **initial condition**, and the line on which the initial condition is specified is then the **initial line**.

The Traveling Wave Equation

The simplest possible form of wave propagation described by the PDE in (26) occurs when $p(x, t, u) = c$ and $q(x, t, u) = 0$, causing the equation to simplify to

traveling wave equation or the advection equation

$$u_t + cu_x = 0 \quad (c = \text{ constant}). \tag{27}$$

This is a linear homogeneous constant coefficient first order PDE that is often known as the **advection equation**.

The classical general solution of (27) can be found by inspection, but for what is to follow it will be more useful if it is obtained by the method of characteristics. Using the characteristic equations (23) and (24) with the new independent variables x and t, we find that the characteristic curves are determined by integrating the equation

$$\frac{dx}{dt} = c \quad \text{to obtain} \quad x = ct + \xi,$$

where the characteristic curve passes through the point $(\xi, 0)$ on the x-axis (the initial line). As $c =$ constant, the characteristics are all parallel straight lines, and the equation of the characteristic through the point $(\xi, 0)$ has the equation

$$x - ct = \xi. \tag{28}$$

The solution $u(x, t)$ along the characteristic curve (line) through the point $(\xi, 0)$ follows by integrating the compatibility equation

$$\frac{du}{dt} = 0 \quad \text{to obtain} \quad u(x, t) = f(\xi).$$

As $u(x, t)$ is constant on a characteristic, the constant value must be equal to the value assigned by the initial condition at the point where the characteristic intersects the x-axis. It follows from this that along the characteristic $x - ct = \xi$ that passes through the point $(\xi, 0)$ we must have $u(x, t) = f(\xi)$. Substituting for ξ shows that the general solution of (27) is

$$u(x, t) = f(x - ct). \tag{29}$$

The derivative $dx/dt = c$ has the dimensions of a speed, so (29) shows that the **profile** of the initial disturbance determined by the function $f(x)$ at the time $t = 0$ is propagated with speed c, without change of shape or scale (size), in the positive x-direction when $c > 0$, and in the negative x-direction when $c < 0$. A wave of this type is called a **traveling wave**, and sometimes a **wave of constant form**. Figure 18.4 shows a typical traveling wave with an initial wave profile in the form of a symmetrical pulse and a propagation speed $c = 2$. The plot illustrates the steady propagation to the right of the initial profile in such a way that at a time $t = t_1$ each point has moved to the right through a distance $2t_1$.

wave profile and a traveling wave

A Typical Linear Constant Coefficient Nonhomogeneous Equation

Let us consider the initial value problem

$$u_t + 3u_x - u = kx, \quad \text{with} \quad u(x, 0) = \sin x \qquad (k = \text{constant}).$$

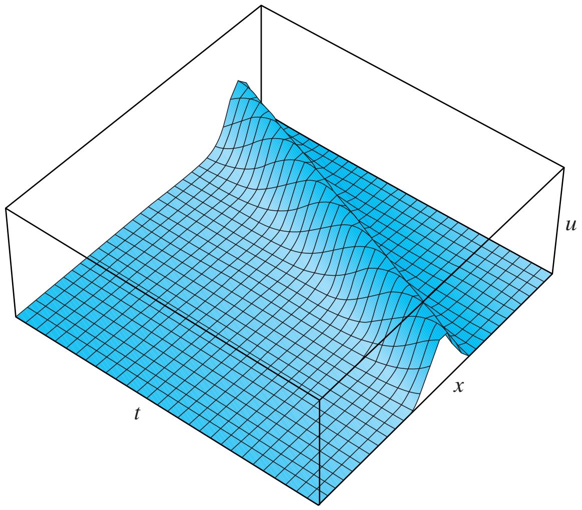


FIGURE 18.4 A traveling wave moving in the positive x-direction with $c = 2$.

traveling wave problems involving linear, semilinear, and quasilinear PDEs

The characteristics determined by integrating $dx/dt = 3$ are $x = 3t + \xi$, where the characteristic intersects the initial line at $(\xi, 0)$. The compatibility condition is $du/dt = u + kx$, but $x = 3t + \xi$ along the characteristic through $(\xi, 0)$, so along this characteristic u is determined by the solution of the ODE

$$\frac{du}{dt} = u + 3kt + k\xi.$$

Solving this linear first order ODE shows that

$$u(x, t) = e^t f(\xi) - k(3t + 3 + \xi),$$

where $f(\xi)$ with f an arbitrary function represents the arbitrary additive integration constant introduced by the integration.

As $\xi = x - 3t$ this solution becomes

$$u(x, t) = e^t f(x - 3t) - k(3 + x).$$

To determine the form of the function f, we now make use of the initial condition $u(x, 0) = \sin x$. Setting $t = 0$ in the expression for $u(x, t)$ and using the initial condition we have

$$\sin x = f(x) - k(3 + x),$$

and so

$$f(x) = \sin x + k(3 + x).$$

Finally, replacing x in $f(x)$ by $x - 3t$ and substituting the result in $u(x, t)$ we arrive at the result

$$u(x, t) = e^t\{\sin(x - 3t) + k(3 + x - 3t)\} - k(3 + x).$$

This expression satisfies the initial condition and the PDE, so it is the required classical solution. Although the speed of propagation of the wave is constant, because $dx/dt = 3$, the wave shape changes from the initial sinusoid as it propagates.

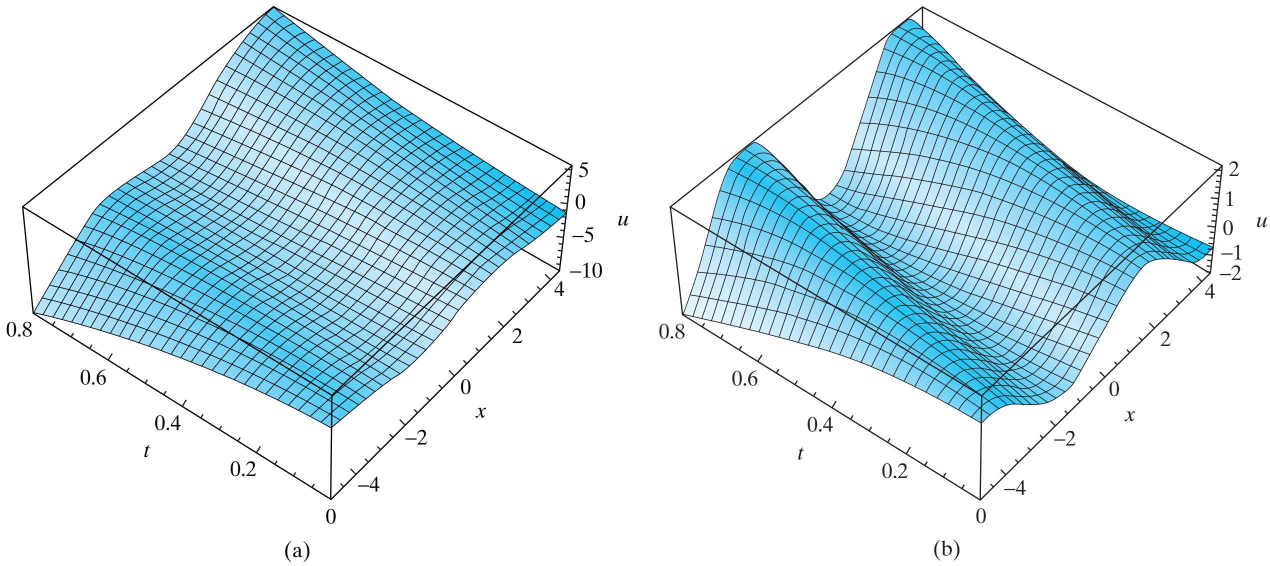


FIGURE 18.5 (a) The solution when $k = 1$; (b) the solution when $k = 0$.

Only when $k = 0$ is the shape of the wave preserved, though *not* its scale, because of the presence of the multiplicative scale factor e^t. Figure 18.5a shows a plot of the solution when $k = 1$ and a plot when $k = 0$ is shown in Fig. 18.5b, in each case for $-5 \le x \le 5$ and $0 \le t \le 0.8$. ■

A Typical Linear Variable Coefficient Nonhomogeneous Equation

The following PDE illustrates the wave propagation properties of a typical linear variable coefficient nonhomogeneous equation. Consider the initial value problem

$$u_t + xu_x + u = 1, \quad \text{with } u(x, 0) = \tanh x.$$

The characteristic curves are determined by integrating the equation

$$\frac{dx}{dt} = x \quad \text{to obtain } x = \xi e^t,$$

where the characteristic curve passes through the point $(\xi, 0)$ on the initial line $t = 0$.

The compatibility condition is

$$\frac{du}{dt} = 1 - u,$$

so when this is integrated along a characteristic curve we find that

$$u = 1 + e^{-t} f(\xi),$$

where f is an arbitrary function of ξ. Substituting for ξ we have

$$u = 1 + e^{-t} f(xe^{-t}).$$

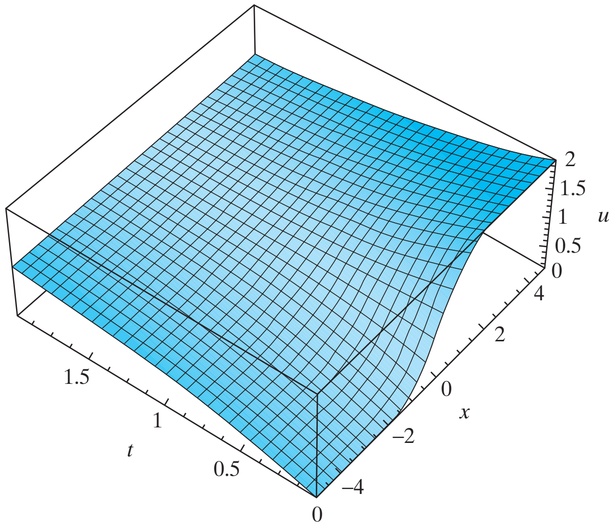


FIGURE 18.6 Decay of the initial condition $u(x, 0) = \tanh x$ to the constant value $u = 1$.

The arbitrary function f must be determined by using the initial condition $u(x, 0) = \tanh x$. Setting $t = 0$ in the preceding expression for u and imposing the initial condition gives

$$\tanh x = 1 + f(x), \quad \text{so that } f(x) = \tanh x - 1.$$

Replacing x in $f(x)$ by xe^{-t} and using the result in the expression for u gives

$$u(x, t) = 1 + e^{-t} \tanh(xe^{-t}).$$

Wave propagation described by this PDE is not at a constant speed, because $dx/dt = x$, nor is its initial shape preserved. Examination of the solution shows that the wave profile changes shape as it propagates, and that after a suitable period of time the profile decays to the constant solution $u(x, t) = 1$, as illustrated in Fig. 18.6. ■

The last examples show that, in general, wave propagation described by first order linear equations that are *not* of the form of (27) describe wave propagation that may or may not preserve the shape of the initial wave profile, but will not preserve the scale as time evolves, so their solutions are not traveling waves.

A Typical Semilinear Equation

The properties of semilinear PDEs can be illustrated by considering the initial value problem

$$u_t + u_x = u^2, \quad \text{with } u(x, 0) = \sin x.$$

The characteristic passing through the point $(\xi, 0)$ in the (x, t)-plane obtained by integrating $dx/dt = 1$ is $x = t + \xi$, and the compatibility condition along this characteristic is $du/dt = u^2$. Integrating the compatibility condition along the characteristic gives

$$-\frac{1}{u} = t + f(\xi),$$

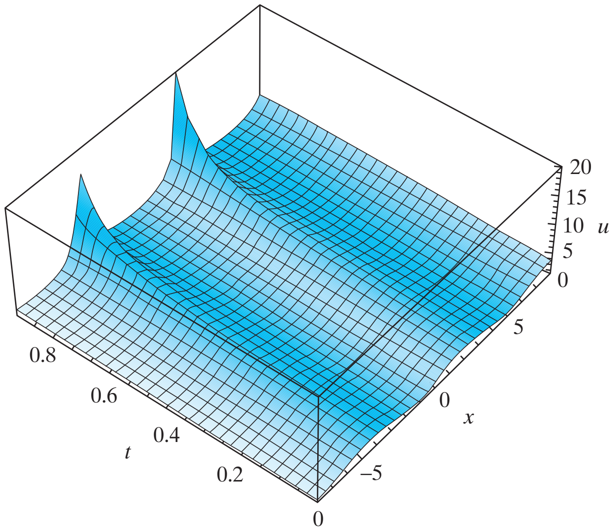


FIGURE 18.7 The evolution of infinite values of $u(x, t)$ as $t \to 1$.

where f is an arbitrary function of ξ. Substituting $\xi = x - t$ into this result, we have

$$u(x, t) = \frac{-1}{t + f(x - t)}.$$

As $u(x, 0) = \sin x$, setting $t = 0$ in $u(x, t)$ and using the initial condition shows that $f(x) = -1/\sin x$, from which it follows that $f(x - t) = -1/\sin(x - t)$. Substituting for $f(x - t)$ in the expression for $u(x, t)$ then gives

$$u(x, t) = \frac{\sin(x - t)}{1 - t\sin(x - t)}.$$

This function satisfies both the initial condition and the PDE, so it is the required classical solution.

Examination of this solution shows that it is only defined in the strip $0 < t < 1$, because only in this strip is the denominator of $u(x, t)$ nonzero. So, unlike linear equations, this semilinear equation has a classical solution for only a finite time, after which for some x the solution becomes infinite. The plot of $u(x, t)$ in Fig. 18.7 shows the development of infinite values of the solution as $t \to 1$. ■

A Typical Quasilinear Equation

The general properties of solutions of the first order quasilinear PDE

$$u_t + p(x, t, u)u_x + q(x, t, u) = 0 \tag{30}$$

can all be illustrated by considering the typical initial value problem

$$\boxed{u_t + f(u)u_x = 0, \quad \text{with } u(x, 0) = g(x),} \tag{31}$$

where f and g are arbitrary functions of their arguments.

The characteristics of (31) are determined by integrating $dx/dt = f(u)$, while the compatibility condition determining the solution u that is valid along a characteristic is seen to be $du/dt = 0$.

The compatibility condition shows that $u =$ constant along a characteristic, with the value of the constant determined by the initial condition at the point of intersection of the characteristic and the initial line. Furthermore, as $u =$ constant along a characteristic, it follows from $dx/dt = f(u)$ that all characteristics will be straight lines, and that the propagation speed $f(u)$ associated with a characteristic is determined by the constant value of u that is transported along it.

Thus, the characteristic through the point $(\xi, 0)$ on the initial line (the x-axis) where the initial condition is $u = g(\xi)$ will have the equation

$$x = \xi + f(g(\xi))t, \quad \text{and along this characteristic } u = g(\xi). \tag{32}$$

Elimination of ξ between these equations, where it appears as a parameter, shows that the solution u of the initial value problem in (31) is determined by the *implicit* relationship

$$u = g\{x - f(u)t\}. \tag{33}$$

To examine the nature of solutions of (31) we must consider the behavior of the characteristic curves (lines in this case), and when doing so we follow the usual convention that the x-axis is taken to be horizontal and the t-axis vertical. Consequently, when drawn in the (x, t)-plane, the gradient of a characteristic curve is $dt/dx = 1/f(u)$.

how solutions of quasilinear PDEs can break down

Let us now suppose that the function $f(u)$ in (31) is a steadily *increasing* function of u. Then the characteristics radiating out from points on the initial line will all fan out, as illustrated in Fig. 18.8a. This shows that the initial value problem (31) will have a unique solution throughout the upper half of the (x, t)-plane, because the solution at any point will be the value of u associated with the characteristic that passes through the point, and the characteristics never intersect. However, if $f(u)$ is a steadily *decreasing* function of u, the characteristics radiating out from points on the initial line will converge, leading to the intersection of characteristics as shown in Fig. 18.8b. When this happens the nature of the solution changes dramatically, because different characteristics transport *different* constant values of u into the upper half of the (x, t)-plane, so the intersection of characteristics corresponds

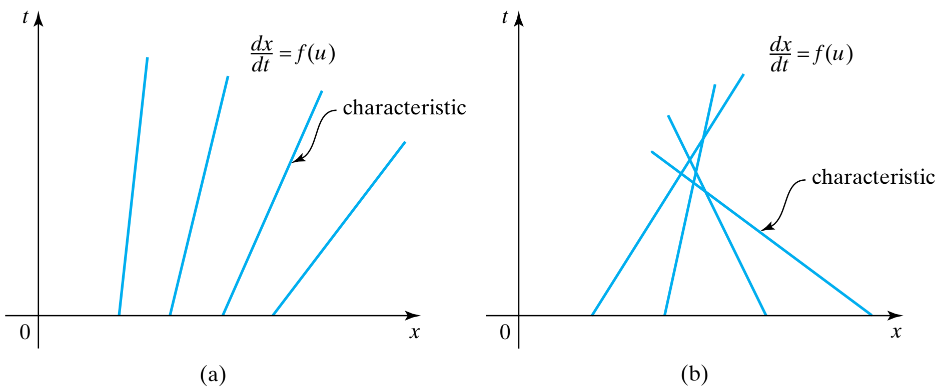


FIGURE 18.8 The influence of $f(u)$ on the behavior of characteristics. (a) $f(u)$ an increasing function of u; (b) $f(u)$ a decreasing function of u.

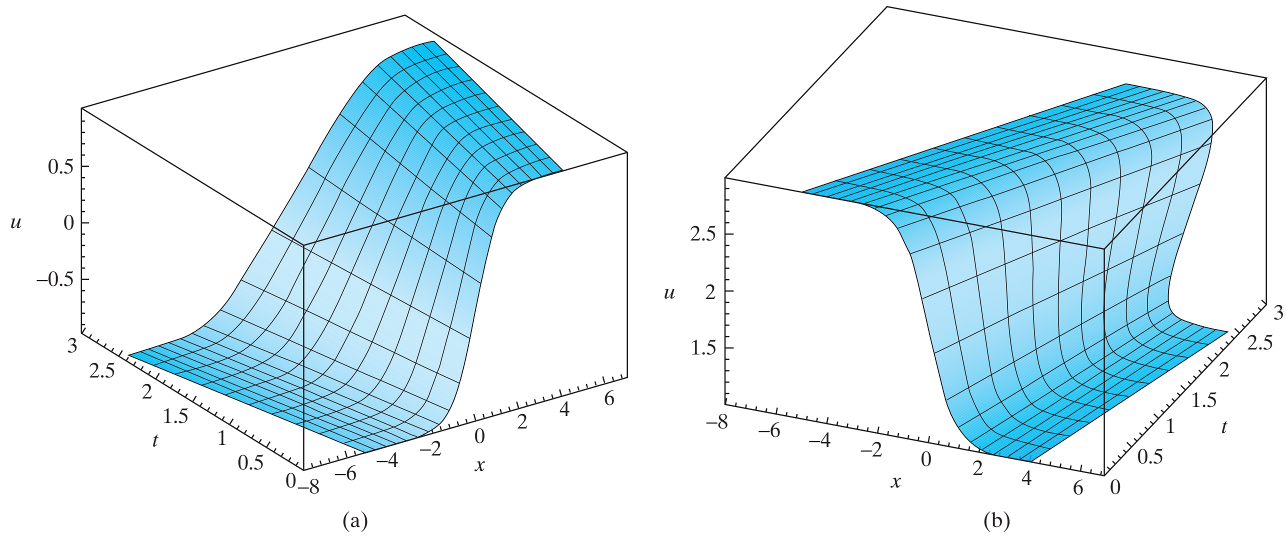


FIGURE 18.9 (a) $f(u)$ an increasing function leading to smoothing because the top of the wave then moves slower than the bottom; (b) $f(u)$ a decreasing function leading to steepening due to the top of the wave moving faster than the bottom.

to the nonuniqueness of the solution of the initial value problem (31) wherever intersection of characteristics occurs. This conclusion is implied by the implicit form of the solution found in (33), because it is known from analysis that a function determined by an implicit relationship need not be unique.

The qualitative properties of waves propagated by a PDE of the form

$$u_t + f(u)u_x = 0$$

can be deduced from the equation $dx/dt = f(u)$ determining the characteristics along which constant initial values of u are transported. To see this, suppose $f(u)$ is an increasing function of u, and consider the wave profile $u(x, t)$. Then if P and Q are adjacent points on a wave profile, with Q to the right of P and $u(Q) > u(P)$, it follows that point Q will propagate faster than point P, causing the wave to become *smoother* as it evolves, as illustrated in Fig. 18.9a. When the converse is true, and $f(u)$ is a decreasing function of u, point P will propagate faster than point Q, causing the wavefront to *steepen*, and eventually this will cause the solution to become nonunique because of the intersection of characteristics, as illustrated in Fig. 18.9b.

Partial differentiation of (33) with respect to x gives

$$\frac{\partial u}{\partial x} = g'\{x - f(u)t\}\left\{1 - f'(u)\frac{\partial u}{\partial x}t\right\},$$

so

$$\frac{\partial u}{\partial x} = \frac{g'\{x - f(u)t\}}{1 + g'\{x - f(u)t\}\, f'(u)t}. \tag{34}$$

This result shows u_x can become infinite at finite time $t = t_c$ if the functions f and g are such that $g'\{x - f(u)t\}f'(u) < 0$ where t_c is the *smallest* time for which $1 + g'\{x - f(u)t\}f'(u)t = 0$. The development of an infinite derivative u_x corresponds to the time when a tangent to the wave profile first becomes vertical, marking the start of the nonuniqueness. This feature can be seen in Fig. 18.9b, where the tangent to the mid-point of the wave profile tends to a vertical position as $t \to 1$. An immediate consequence of this is that when characteristics converge, a classical solution can only exist for a finite time in the strip $0 < t < t_c$ in the (x, t)-plane.

Solutions of initial value problems for the more general first order quasilinear PDE in (30) exhibit the same general properties as those of (31). As typical functions $p(x, t, u)$ in (30) and $f(u)$ in (31) will have domains where they are increasing functions of u and others where they are decreasing functions, in general classical solutions of first order quasilinear PDEs can only exist for a finite time. The next section examines how the concept of a solution can be extended to allow the solution of some PDEs to be generalized so that a solution can be extended beyond the time t_c.

EXAMPLE 18.5 Solve the initial value problem

$$u_t + (1 + u)u_x + u = 0, \quad \text{given that} \quad u(x, 0) = 1 + x.$$

Solution This PDE is quasilinear because of the product term uu_x. The characteristic curves are obtained by integrating $dx/dt = 1 + u$, and the compatibility condition determining u along a characteristic is $du/dt = -u$.

Let the solution along the characteristic through the point $(\xi, 0)$ on the initial line be $u = g(\xi)$, then integration of the compatibility condition gives

$$\ln u = -t + \ln g(\xi), \quad \text{and so} \quad u = g(\xi)e^{-t}.$$

From this result and the initial condition at $(\xi, 0)$ we have $g(\xi) = 1 + \xi$, so the solution can be written

$$u = (1 + \xi)e^{-t}.$$

Substitution of this result into the equation determining the characteristic curves gives

$$\frac{dx}{dt} = 1 + (1 + \xi)e^{-t},$$

and so

$$\int_\xi^x ds = \int_0^t [1 + (1 + \xi)e^{-\tau}]d\tau,$$

where s and τ are dummy variables. After integration this becomes

$$x = \xi(2 - e^{-t}) + t + 1 - e^{-t},$$

from which it follows that

$$1 + \xi = \frac{1 + x - t}{2 - e^{-t}}.$$

Finally, using this result to eliminate ξ from the expression for u, we find that

$$u(x, t) = \left(\frac{1 + x - t}{2 - e^{-t}}\right)e^{-t}.$$

This function satisfies the initial condition and the original PDE, so it is the required classical solution. As the denominator does not vanish for $t > 0$, this is the classical solution for the initial value problem for $t > 0$. ■

More information on the method of characteristics, including applications, can be found in references [7.1], [7.4], [7.6], [7.8], [7.11], [7.12], and [7.20].

Summary

The concept of wave propagation was introduced and related to the method of characteristics. Each characteristic curve was seen to transport the initial condition appropriate to the characteristic according to the ODE determining the evolution of the solution along the curve. It was shown how homogeneous linear first order PDEs can have traveling wave solutions where the shape of the wave remains unchanged as it propagates with time. However, the introduction of nonlinearity was seen to make traveling wave solutions impossible, and in certain cases to lead to the solution becoming nonunique after a finite time.

EXERCISES 18.3

Solve the following initial value problems.

1. $2u_t + 4u_x = 3u$, given that $u(x, 0) = \sin 2x$.
2. $u_t - 2u_x = x$, given that $u(x, 0) = x^2$.
3. $u_t - 3u_x = 2u + 1$, given that $u(x, 0) = \frac{1}{2}\cos x$.
4. $u_t - u_x = u + \sin x$, given that $u(x, 0) = 1$.
5. $u_t - 4u_x = 3x$, given that $u(x, 0) = e^x$.
6. $u_t + 2u_x = 2u + x$, given that $u(x, 0) = x$.
7. $u_t - 3xu_x + 2u = x$, given that $u(x, 0) = x$.
8. $u_t + 3xu_x - 2u = 4$, given that $u(x, 0) = x$.
9. $u_t - 3xu_x + 2u = x$, given that $u(x, 0) = 3x$.
10. $(1 + t^2)u_t + u_x = (1 + t^2)(u - 1)$, given that $u(x, 0) = \sinh x$.
11. $3u_t - 9xu_x + 6u = x$, given that $u(x, 0) = x$.
12. $u_t + e^{2t}u_x = u + x$, given that $u(x, 0) = 1$.
13. $u_t + u_x = 2u^2$, given that $u(x, 0) = \cos x$.
14. $u_t + 4xu_x = u^2$, given that $u(x, 0) = \sinh x$.
15. $u_t + 2uu_x - u = 0$, given that $u(x, 0) = -2x$.
16. $u_t + 2uu_x + 2u = 0$, given that $u(x, 0) = 3x$.
17. $u_t - 3uu_x + 4u = 0$, given that $u(x, 0) = 1 + x$.
18. $u_t + tuu_x - u = 0$, given that $u(x, 0) = 2x$.
19. $u_t + (1 + t)uu_x - \left(\dfrac{1}{1+t}\right)u = 0$, given that $u(x, 0) = 3x - 1$.
20. $u_t + uu_x - \left(\dfrac{1}{1+t}\right)u = 0$, given that $u(x, 0) = 1 - x$.

18.4 Generalizing Solutions: Conservation Laws and Shocks

In many physical situations a commonly occurring feature of wave propagation is the evolution of smooth solutions of PDEs to a point where their nature changes, and jump discontinuities occur and propagate in a manner quite different from the smooth solution. This happens in fluid and solid mechanics, in magnetohydrodynamics, and elsewhere when the governing PDEs are quasilinear and describe wave propagation.

generalizing solutions

The propagation of discontinuities in otherwise continuous and differentiable solutions represents an extension of the concept of a solution that has been used thus far. This is because although the solution on either side of the discontinuity satisfies the original PDE, the solution is not a classical solution since it is not differentiable at a jump discontinuity. In high-speed gas dynamics, and in elastic

materials that behave nonlinearly, discontinuous solutions of this type are called **shock waves**.

Jump discontinuities can also develop and propagate in water, as can be seen in estuaries subject to suitable tidal conditions, where a mass of water across which there is a large and abrupt change of level can propagate in a stable manner for a considerable distance. A steplike disturbance of this type in water is called a tidal **bore**, and when the effects of viscosity and turbulence are neglected the situation can be approximated mathematically by a jump discontinuity in the water height.

Behavior of this type was suggested in the last section where it was seen that classical solutions of initial value problems for first order quasilinear equations may only exist for a finite time until the solution becomes nondifferentiable. This suggests that a possible generalization of a classical solution $u(x, t)$ could involve a function that is differentiable and satisfies a PDE on either side of a moving point $x = \sigma(t)$ inside a fixed interval $x_1 \leq x \leq x_2$, but that across the moving point the solution is discontinuous and experiences a finite jump. Let us see how such a generalization of a solution can be obtained, and in the process examine some of its properties and how it depends fundamentally on the notion of a conservation law.

The fundamental idea that will be used to extend the notion of a classical solution is most easily understood by considering the simple PDE

$$u_t + uu_x = 0, \tag{35}$$

which is a special case of (31) with $f(u) = u$. As $uu_x = \frac{\partial}{\partial x}(\frac{1}{2}u^2)$, the PDE in (35) can be written

$$u_t + \left(\frac{1}{2}u^2\right)_x = 0. \tag{36}$$

To allow for a discontinuity we will use an integral representation of (36), because although the derivative of $u(x, t)$ is not defined at a point where the function is discontinuous, its integral over an interval $x_1 \leq x \leq x_2$ containing the discontinuity is well defined. Let us now attempt to generalize the concept of a solution of (36) to allow for a situation where $u(x, t)$ satisfies the PDE to the left and right of a moving interior point $x = \sigma(t)$ in the interval $x_1 \leq x \leq x_2$, but across which it is discontinuous, with $u = u_L$ at the point $x = \sigma(t)_L$ to the immediate left of $x = \sigma(t)$ and $u = u_R$ at the point $x = \sigma(t)_R$ to the immediate right, with $u_L \neq u_R$.

Integrating (36) over the interval $x_1 \leq x \leq x_2$ gives

$$\int_{x_1}^{x_2} \frac{\partial u}{\partial t} dx + \int_{x_1}^{x_2} \left(\frac{1}{2}u^2\right)_x dx = 0. \tag{37}$$

Provided u is differentiable with respect to t, the time derivative can be taken outside the first integral in (37), which then becomes

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x, t)dx + \int_{x_1}^{x_2} \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right)dx = 0. \tag{38}$$

An application of the fundamental theorem of integral calculus to the second integral leads to the result

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)dx + \frac{1}{2}\{u^2(x_2,t) - u^2(x_1,t)\} = 0. \tag{39}$$

To develop this result further by allowing for the discontinuity in $u(x,t)$ across $x = \sigma(t)$, we now rewrite (39) as

$$\frac{d}{dt}\int_{x_1}^{\sigma(t)_L} u(x,t)dx + \frac{d}{dt}\int_{\sigma(t)_R}^{x_2} u(x,t)dx = \frac{1}{2}\{u^2(x_1,t) - u^2(x_2,t)\}. \tag{40}$$

conservation law in integral form

This result is a **conservation law** in **integral form** for the quantity represented by $u(x,t)$. The term on the left is the rate of change of the amount of $u(x,t)$ in the interval $x_1 \le x \le x_2$, and the term on the right represents the difference between the amount of $u(x,t)$ entering through $x = x_1$ and leaving through $x = x_2$.

If Leibniz' theorem (Theorem 1.5) for the differentiation of a definite integral with respect to a parameter is applied to the term on the left of this equation, we find that

$$\int_{x_1}^{\sigma(t)_L} u_t(x,t)dx + \int_{\sigma(t)_L}^{x_2} u_t(x,t)dx + \frac{d\sigma}{dt}(u_L - u_R) = \frac{1}{2}\{u^2(x_1,t) - u^2(x_2,t)\}. \tag{41}$$

Letting $x_1 \to \sigma(t)_L$ and $x_2 \to \sigma(t)_R$, when $u(x_1,t) \to u_L$ and $u(x_2,t) \to u_R$, simplifies this result to

$$\frac{d\sigma}{dt}(u_L - u_R) = \frac{1}{2}(u_L^2 - u_R^2), \tag{42}$$

because the boundedness of u_t causes the two integrals to vanish in the limit as their intervals of integration tend to zero.

If we set $s = d\sigma/dt$, and introduce the notation $[[\alpha]] = \alpha_L - \alpha_R$, the **jump condition** experienced by a discontinuous solution of this PDE across the discontinuity at $x = \sigma(t)$ becomes

$$s[[u]] = \frac{1}{2}[[u^2]]. \tag{43}$$

In terms of u_L and u_R this can be written

$$s(u_L - u_R) = \frac{1}{2}(u_L^2 - u_R^2), \tag{44}$$

so the speed of propagation of the discontinuity

$$s = \frac{1}{2}(u_L + u_R). \tag{45}$$

shock waves and the Riemann problem

A discontinuity across $x = \sigma(t)$ is called a **shock wave**, or simply a **shock**, when it arises because of the intersection of characteristics.

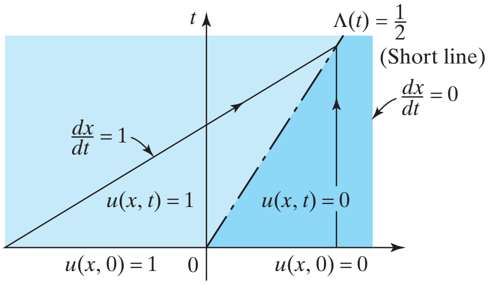


FIGURE 18.10 Characteristics in Riemann problem (I) converge to produce a discontinuous generalized solution that forms a propagating shock wave.

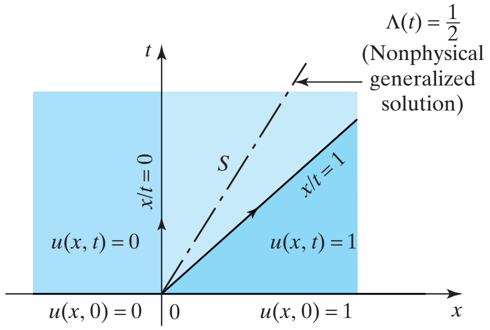


FIGURE 18.11 A mathematically permissible but nonphysical discontinuous solution in S for Riemann problem (II) that is *not* produced by the intersection of characteristics. The two constant solutions to the left and right of region S are joined continuously in a physically realistic manner by a centered simple wave in S.

To illustrate some of the properties of this extension of a classical solution, we now consider two special piecewise continuous initial value problems for (35) called **Riemann problems**.

Riemann problem (I): Solve the initial value problem

$$u_t + uu_x = 0, \quad \text{with} \quad u(x,0) = \begin{cases} 1, & x < 0 \\ 0, & x > 0, \end{cases} \tag{46}$$

where the initial condition is piecewise constant and decreases as x increases.

From (45) the speed of propagation of the discontinuity initiated by the discontinuity in the initial data is seen to be $s = \frac{1}{2}$. Figure 18.10 shows that this propagating discontinuity is a *shock*, because characteristics converge onto the discontinuity line from both the left and right. In gas dynamics a discontinuity of this type models an ideal shock wave in supersonic flow across which there is a sudden change of pressure, which in supersonic flight causes the sonic boom as an aircraft flies past.

Riemann problem (II): Solve the initial value problem

$$u_t + uu_x = 0, \quad \text{with} \quad u(x,0) = \begin{cases} 0, & x < 0 \\ 1, & x > 0, \end{cases} \tag{47}$$

where the initial condition is piecewise constant and increases as x increases.

In this problem the speed of propagation of the discontinuity is again $s = \frac{1}{2}$, but Fig. 18.11 shows that the discontinuity cannot be a shock because no characteristics converge onto the line along which the discontinuity is propagated. In applications, a discontinuous solution of this type is a mathematical solution but *not* a physically realizable one, as was the case in Riemann problem (I). This illustrates the fact that a consequence of extending a classical solution to permit discontinuous solutions can be to introduce nonphysical solutions that must be rejected when they do not arise because of the intersection of characteristics.

a mathematical solution that is nonphysical

To examine Riemann problem (II) in more detail, having rejected its discontinuous generalized solution as not physically realizable, we need to consider how a differentiable solution can be found in the wedge-shaped region S in Fig. 18.11. For a differentiable solution to exist in S it is necessary that the region be covered by a family of characteristics that at the left and right extremes of S coincide with the characteristics bounding the adjacent regions where $u(x, t)$ is constant. This can be achieved by straight line characteristics (rays) emanating from the origin O, the equation of which can be written $\zeta = x/t$ with $0 \leq \zeta \leq 1$, because then the rays at the edges of S coincide with the characteristics bounding the constant state regions.

Let us now try to find a solution of (47) in region S of the form $u(x, t) = U(\zeta)$, where $\zeta = x/t$. Then, as $u_t = U'(\zeta)\partial\zeta/\partial t = -(x/t^2)U'(\zeta)$ and $u_x = U'(\zeta)\partial\zeta/\partial x = (1/t)U'(\zeta)$, substitution into (35) followed by the cancellation of t and $U'(\zeta)$, neither of which is zero, shows that

$$U(\zeta) = u(x, t) = x/t. \tag{48}$$

This is the required solution of Riemann problem (II) in S. The solution $u(x, t)$ in S is constant along every characteristic issuing out from the origin, and at the extremes of S these characteristics coincide with the characteristics bounding the constant solutions to the left and right of S. This solution in S resolves the initial discontinuity immediately and joins the constant solutions to the left and right of S in a continuous manner. A solution of this type is called a **centered simple wave** with its center located at the origin 0. This is a *generalized solution* because of the discontinuity in derivatives across the characteristics that bound S. In applications, a centered simple wave resolves discontinuous initial conditions that do not give rise to the intersection of characteristics, and in Riemann problem (II) the non-physical discontinuous generalized solution that is also possible must be rejected and replaced by the physically realizable centered simple wave.

centered simple wave

A proper examination of shock waves, centered simple waves, and simple waves of a more general type is beyond this brief introduction, as is a discussion of a different form of generalization of a solution called a **weak solution**. Nevertheless, the extension of a classical solution outlined here to include shock solutions has many important practical consequences, as, for example, in fluid mechanics, solid mechanics, and electromagnetic theory. In three space dimensions and time, these ideas are used to examine shock waves produced by aircraft in supersonic flight, and the bow shock wave produced by the Space Shuttle during its reentry into the atmosphere.

A classical account of shock waves in gases can be found in reference [7.4]. References [7.9] and [7.13] consider the generalization of differentiable solutions of PDEs to allow for discontinuous solutions; see also reference [7.20]. Reference [7.13] also covers in considerable detail various types of reaction–diffusion problems. A useful and elementary introduction to the mathematical theory of waves of several different types is to be found in reference [7.8]; reference [7.10] develops the mathematical theory of PDEs in considerable detail. A standard reference to various types of wave propagation problem is to be found in reference [7.18].

Summary

It was shown how, when a first order PDE describing a conservation law is written in integral form, it is possible to extend the classical concept of a differentiable solution by incorporating discontinuous solutions called shocks. This becomes necessary in order to extend the concept of a solution to take into account the situation when the classical solution becomes nonunique because of the intersection of characteristics, causing the

solution to become nondifferentiable. It was seen that this generalization of a solution can give rise to more than one shock solution. In physical situations, such as gas dynamics, only one of these shock solutions is possible, so some selection principle must be introduced to allow the physically realizable solution to be distinguished from among the mathematically possible ones.

EXERCISES 18.4

1. Find the jump condition that must be satisfied by a shock solution of

$$u_t + u^n u_x = 0 \quad \text{for } n = 1, 2, \ldots.$$

2. Given that the differential equation

$$u_t + f(u)u_x = 0$$

has a discontinuous solution and that $f(u)$ is a continuous function of u, find the jump condition that must be satisfied by its shock solution.

3. Given the two Riemann problems for the equation

$$u_t + u^2 u_x = 0,$$

determined by (a) $u(x,0) = \begin{cases} 1, & x < 0 \\ 2, & x > 0 \end{cases}$ and (b) $u(x,0) = \begin{cases} 3, & x < 0 \\ 1, & x > 0 \end{cases}$, find which problem has a shock solution and determine its speed of propagation.

4. Show that the Riemann problem

$$u_t + u^3 u_x = 0 \quad \text{with} \quad u(x,0) = \begin{cases} 1, & x < 0 \\ 2, & x > 0 \end{cases}$$

has a centered simple wave solution located at the origin. By setting $\zeta = x/t$, $u(x,t) = U(\zeta)$ and substituting into the differential equation, find the analytical solution for the centered simple wave and determine the region in the (x,t)-plane occupied by the simple wave solution.

5.* Show that the Riemann problem

$$u_t + uu_x = 0 \quad \text{with} \quad u(x,0) = \begin{cases} 0, & x < 2 \\ 1, & x > 2 \end{cases}$$

has a centered simple wave solution. Generalize the approach suggested in Exercise 4 to find the analytical solution for the centered simple wave, stating the region in the (x,t)-plane occupied by the centered simple wave solution.

6.* The compound Riemann problem

$$u_t + uu_x = 0 \quad \text{with} \quad u(x,0) = \begin{cases} 1, & x < 0 \\ 2, & 0 < x < 2 \\ 0, & x > 2 \end{cases}$$

describes a solution that starts with both a centered simple wave and a shock located at different points on the initial line. By considering the path of the shock and the boundary of the centered simple wave, determine the time at which the simple wave and the shock first meet.

18.5 The Three Fundamental Types of Linear Second Order PDE

We now show how the three most important types of linear second order PDEs can be derived from some representative physical problems. The equations are classified as being of hyperbolic, parabolic, or elliptic type, and the basis of this system of classification will be developed in the next section.

Vibrating Strings and Plates

vibrating strings and plates and the wave equation

Let us consider a uniform stretched linearly elastic string under a tension T that is displaced from its equilibrium position and then released. This could, for example, represent the response of a plucked violin string. To derive the PDE governing the motion of the string after its release we must examine the forces acting on an element PQ of the string at time t when it has been displaced through a small

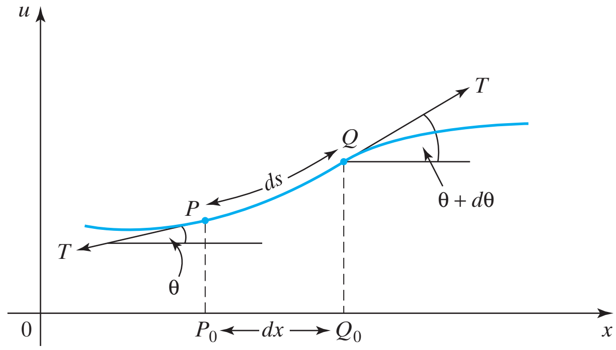


FIGURE 18.12 A transverse displacement of element PQ of a stretched string.

distance in the u-direction transverse to its equilibrium position along the x-axis. Figure 18.12 shows a typical element PQ when in its displaced position.

The element of arc length ds along the string when the displacement is $u(x, t)$ is given by $ds = \sqrt{(1 + u_x^2)}dx$. As the displacement u is small, the term u_x^2 is small relative to 1, so to this order of approximation $ds \approx dx$. In a linearly elastic string the tension is proportional to the extension of the string, so as $ds \approx dx$ we may assume that the string tension T remains constant as long as the transverse displacement is small.

In the equilibrium condition let the element PQ lie along the x-axis between the points P_0 and Q_0, where the length PQ is ds and the length P_0Q_0 is dx. The equation of motion of the element is obtained by equating the forces acting on the element due to the tension T (gravity is neglected) and the rate of change of momentum of the element in the u-direction. As the string is uniform, the mass of the element PQ is $dm = \rho ds$, where ρ, called the **line density** of the string, is the mass per unit length of the string. The momentum of the element in the u-direction is $\rho ds\, u_t$, so its rate of change of momentum in the direction is $\rho ds\, u_{tt}$. As T is considered to be constant, the force acting on the element is simply the difference in the components of the tension normal to the x-axis at each of its ends due to the change in inclination of the string from an angle θ at P to an angle $\theta + d\theta$ at Q. The resultant force acting on the element is thus

$$T \sin(\theta + d\theta) - T \sin\theta = T \sin\theta \cos d\theta + T \cos\theta \sin d\theta - T \sin\theta.$$

As $d\theta$ is small we may replace $\cos d\theta$ by 1 and $\sin d\theta$ by $d\theta$, as a result of which the transverse force acting on the string can be approximated by $T \cos\theta d\theta$. Finally, equating the resultant force and the rate of change of momentum in the u-direction shows that when $d\theta$ and the transverse displacements are small the equation of motion is

$$T \cos\theta d\theta = \rho ds\, u_{tt}.$$

To eliminate θ we now use the fact that $\tan\theta = u_x$, from which it follows by differentiation with respect to x that $\sec^2\theta\, d\theta/dx = u_{xx}$, and so $\sec^2\theta\, d\theta = u_{xx}dx$. Multiplying this by $T\cos^3\theta$, substituting into the preceding result, and using the fact that in the limit as $dx \to 0$ we have $dx/ds = \cos\theta$ leads to the result

$$\rho u_{tt} = T \cos^4\theta\, u_{xx}.$$

As $\tan\theta = u_x$ and $\sec^2\theta = 1 + \tan^2\theta$, we see that

$$\cos^2\theta = 1/\{1 + (u_x)^2\},$$

so the equation of motion becomes

$$u_{tt} = c^2\{1 + (u_x)^2\}^{-2}u_{xx},$$

where $c^2 = T/\rho$. This second order partial differential equation governing the motion of the string is quasilinear, but if the transverse displacement is sufficiently small the term $(u_x)^2$ can be neglected, the *linearized* one-dimensional form of the equation of motion becomes

the wave equation is the prototype second order hyperbolic PDE

$$u_{tt} = c^2u_{xx}. \tag{49}$$

This is a linear second order PDE of *hyperbolic* type called the **one-dimensional wave equation**, and it is one of the three fundamentally different classes of second order PDE.

Vibrations of membranes can be treated in a similar fashion to vibrating strings. Figure 18.13 shows a vibrating rectangular element $ABCD$ of a thin uniform membrane with its sides of lengths dx and dy parallel to the x- and y-axes displaced a small amount in the u-direction normal to its equilibrium position in the (x, y)-plane (the plane $u = 0$). If L is a line of unit length drawn in the membrane, the **tension** T in the membrane is defined as the force exerted on L by the material on one side of the line. The tension will be said to be *uniform* when T is independent of the direction of L and its location in the membrane.

Reasoning as in the case of the vibrating string, and considering a membrane with a *uniform* tension T, we see that the resultant of the forces Tdx normal to the boundaries AB and CD of the element is $(Tdx)(u_{yy}dy)$ and, similarly, the resultant of the forces Tdy normal to the boundaries AD and BC of the element is $(Tdy)(u_{xx}dx)$. If the mass per unit area of the membrane ρ, called its **area density**, is constant, the momentum of the element in the u direction is $\rho dxdyu_t$, so its rate of change of momentum in that direction is $\rho dxdyu_{tt}$. Equating the forces acting to the rate of change of momentum and proceeding to the limit as $dx \to 0$ and

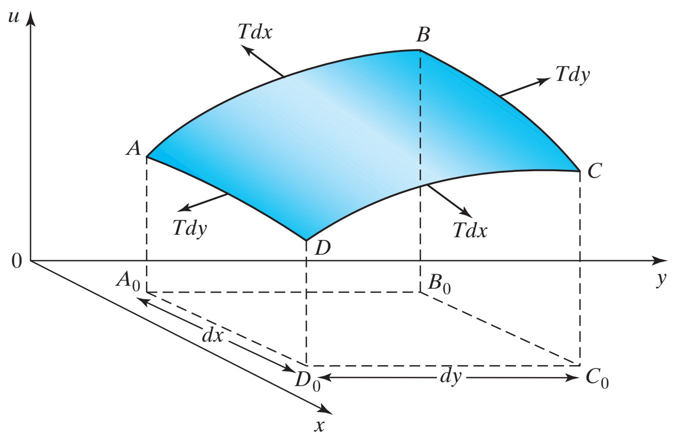


FIGURE 18.13 An element of a uniform vibrating membrane with tension T.

$dy \to 0$, we find that the PDE describing the vibrations is

$$\rho u_{tt} = T\{u_{xx} + u_{yy}\},$$

and after we set $c^2 = T/\rho$ this becomes

$$u_{tt} = c^2\{u_{xx} + u_{yy}\}. \tag{50}$$

This linear second order PDE, which is also of *hyperbolic* type, is called the **two-dimensional wave equation**. Notice that the one-dimensional and two-dimensional wave equations have *second order* partial derivatives with respect to both the *time* and the *space variables* involved.

The Heat (Diffusion) Equation

We now derive the **heat equation**, also known as the **diffusion equation**, that describes the flow of heat through a heat-conducting solid material. The derivation is based on the experimentally observed fact that heat flows in the direction of decreasing temperature, and on the assumption that the rate of heat flow **j** at any point P in the body is given by **Fourier's law**

$$\boxed{\mathbf{j} = -K \operatorname{grad} T,} \tag{51}$$

where $T(x, y, z, t)$ is the temperature at any point P in the material at time t, and K, called the **thermal conductivity** of the material, is a physical property that is usually taken to be a constant.

If V is an arbitrary volume in the solid bounded by a surface S, the quantity of heat leaving V in unit time is given by the surface integral

$$\int_S \mathbf{j} \cdot \mathbf{n} dS, \tag{52}$$

where $\mathbf{n}$ is the *outward* drawn unit normal to S. If we substitute for $\mathbf{j}$ in (52) and allow K to be a function of position, an application of the divergence theorem to this integral gives

$$\int_S \mathbf{j} \cdot \mathbf{n} dS = -\int_V \operatorname{div}(K \operatorname{grad} T) dV.$$

However, $\operatorname{div}(K \operatorname{grad} T) = K\Delta T + \operatorname{grad} K \cdot \operatorname{grad} T$, so the preceding expression becomes

$$\int_S \mathbf{j} \cdot \mathbf{n} dS = -\int_V (K\Delta T + \operatorname{grad} K \cdot \operatorname{grad} T) dV. \tag{53}$$

If the density of the material is ρ and its specific heat is c, the amount of heat in an element of volume dV is given by $c\rho T dV$, where both ρ and c can be functions of position. Integration of $c\rho T dV$ over V shows that the total amount of heat Q in V must be

$$Q = \int_V \rho c T dV.$$

As V is a fixed arbitrary volume in the solid, differentiating this result with respect to the time t shows that the rate at which Q decreases with respect to

time is

$$-Q_t = -\int_V \frac{\partial}{\partial t}(\rho c T)dV. \tag{54}$$

Equating (53) and (54) and combining the integrals gives

$$\int_V \left\{ \frac{\partial}{\partial t}(\rho c T) - K\Delta T - \text{grad } K \cdot \text{grad } T \right\} dV = 0. \tag{55}$$

This result must be true for all arbitrary volumes V, but this can only be possible if the integrand of (55) is identically zero, so the PDE determining the flow of heat when expressed in terms of the temperature T is

the heat or diffusion equation is the prototype second order parabolic PDE

$$\frac{\partial}{\partial t}(\rho c T) = K\Delta T + \text{grad } K \cdot \text{grad } T. \tag{56}$$

This PDE is a linear variable coefficient second order PDE for the temperature distribution throughout the solid, and in general its independent variables are three space variables and time. When, as is usually the case, the conductivity K, the density ρ, and the specific heat c are taken to be constants, the linear second order PDE in (56), which is an equation of *parabolic* type, reduces to

$$\rho c T_t = K\Delta T, \tag{57}$$

heat conduction and diffusion

called the **heat conduction equation**, or simply the **heat equation**. The constant κ^2, where $\kappa^2 = K/(\rho c)$, is called the **diffusivity** of the material, so in terms of the diffusivity, (57) becomes

$$T_t = \kappa^2 \Delta T. \tag{58}$$

Values of the diffusivity κ^2 for some common materials, in c.g.s. units and degrees Celsius, are steel 0.12, copper 1.14, aluminum 0.86, silver 1.71, glass 0.006, and concrete 0.004.

Notice that the heat equation that is of *parabolic type* involves a *first order* partial derivative with respect to time and *second order* partial derivatives with respect to the space variables involved.

An equation of the form (58) also describes the diffusion process caused by an imbalance of concentration of a substance diffusing through material, and for this reason (58) is also known as the **diffusion equation**. A typical diffusion process involves the passage of a chemical with concentration k_1 present in a liquid or gas through a membrane to a liquid or gas on the other side of the membrane where the concentration is k_2 with $k_1 > k_2$.

Diffusion is used in many ways for the concentration of chemicals, and it occurs naturally in plants where nutrients obtained from the soil are passed through the plant by diffusion through plant membranes.

The Laplace Equation

the Laplace equation is the prototype second order elliptic PDE

The **Laplace equation** characterizes a large group of physical problems that are *independent* of the time, and for this reason they are usually called **steady state problems**. An obvious example is provided by the heat equation in (58), because if a heat transfer process attains a steady state the time derivative T_t vanishes and the heat equation reduces to the Laplace equation $\Delta T = 0$ that is the simplest PDE of *elliptic* type. Some typical two-dimensional steady state temperature distributions have already been obtained in Section 17.2 as applications of conformal transformation techniques, where it was also shown that Laplace's equation governs the velocity potential of the steady fluid flow of an incompressible, irrotational, and inviscid fluid.

Other physical situations that give rise to Laplace's equation can be found in the study of steady state electromagnetic fields. When the field exists in an isotropic medium with dielectric constant ε, permeability μ, and charge distribution density ρ, the electric vector $\mathbf{E}$, the magnetic vector $\mathbf{H}$, and the current $\mathbf{j}$ are related by the **Maxwell equations**

the Maxwell equations of electromagnetic theory

$$\begin{aligned}\operatorname{curl}\mathbf{H} &= \mathbf{j} + \varepsilon\frac{\partial}{\partial t}\mathbf{E}\\ \operatorname{curl}\mathbf{E} &= -\mu\frac{\partial}{\partial t}\mathbf{H}\\ \operatorname{div}\mathbf{H} &= 0\\ \operatorname{div}\mathbf{E} &= \rho/\varepsilon.\end{aligned} \tag{59}$$

In **electrostatics** there is no change with respect to time of the electric vector $\mathbf{E}$, so the time derivative $\mathbf{E}_t$ vanishes, and in an uncharged region ($\rho = 0$) Maxwell's equations reduce to

$$\operatorname{div}\mathbf{E} = 0 \quad \text{and} \quad \operatorname{curl}\mathbf{E} = \mathbf{0}.$$

This pair of equations can be satisfied by introducing a scalar **electric potential** ϕ such that $\mathbf{E} = \operatorname{grad}\phi$, because then $\operatorname{curl}\mathbf{E} = \operatorname{curl}(\operatorname{grad}\phi) = \mathbf{0}$, so

$$\operatorname{div}\mathbf{E} = \operatorname{div}(\operatorname{grad}\phi) = 0 \quad \text{and so} \quad \Delta\phi = 0. \tag{60}$$

This has shown that the electrostatic potential distribution ϕ is a solution of the Laplace equation, and that the electric field vector can be found from ϕ by using $\mathbf{E} = \operatorname{grad}\phi$. Various electrostatic potential distributions were found in Section 17.2 by means of conformal transformations.

electrostatics and magnetostatics

A similar situation occurs in **magnetostatics**, because if the medium is nonconducting $\mathbf{j} = \mathbf{0}$, so the Maxwell equations reduce to

$$\operatorname{div}\mathbf{H} = 0 \quad \text{and} \quad \operatorname{curl}\mathbf{H} = \mathbf{0}.$$

This time a **magnetic potential** ϕ can be introduced by setting $\mathbf{H} = \operatorname{grad}\phi$, and then the magnetic potential is seen to be a solution of the Laplace equation $\Delta\phi = 0$.

An important physical problem that gives rise to the Laplace equation in three dimensions is the gravitational potential $\phi(x, y, z)$. The mathematics of gravitational potentials is closely related to the cases considered above, but before we proceed further, some definitions are necessary.

A **force field** in a region D of space exerts a force $\mathbf{F}$ on a material solid particle at a point (x, y, z) in D, where

$$\mathbf{F} = F_1(x, y, z)\mathbf{i} + F_2(x, y, z)\mathbf{j} + F_3(x, y, z)\mathbf{k}.$$

It may happen that the force is proportional to the mass m of the particle, as occurs in the earth's gravitational field, where the constant of proportionality between the mass of the particle and its weight is g, the acceleration due to gravity.

force fields and lines of force

A curve in a force field with the property that at each point on the curve the tangent to the curve is parallel to the direction of the force is called a **line of force**. If the vector element along the line of force is $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$, the lines of force are determined by the equations

$$\frac{dx}{F_1} = \frac{dy}{F_2} = \frac{dz}{F_3}. \tag{61}$$

When a particle moves in a force field from A to B along a path AB, the work W done by the action of the field on the particle is given by the line integral

$$W_{AB} = \int_{AB} (F_1 dx + F_2 dy + F_3 dz).$$

In general, the work W_{AB} will depend not only on A and B, but also on the path taken from A to B. A **potential field** is a force field in which the work done by the force depends only on the points A and B, and not on the path joining them. Consequently, a field is a *potential field* if the work done along every loop joining A to itself is zero. It is for this reason that potential fields are also called **conservative fields**, because work done by the force on a particle moving away from a point is returned if the particle arrives back at its starting point.

potentials and conservative fields

Consider the two arbitrary paths APB and AQB shown in Fig. 18.14a. Then in a potential field $W_{APB} + W_{BQA} = 0$, so $W_{BQA} = -W_{APB}$. Now let A in Fig. 18.14b be a fixed point (x_0, y_0, z_0), B be a general point (x, y, z), and C be a point (x^*, y^*, z^*). Then if $W_{AB} = \phi(x, y, z)$ is the work done moving from A to B, in a

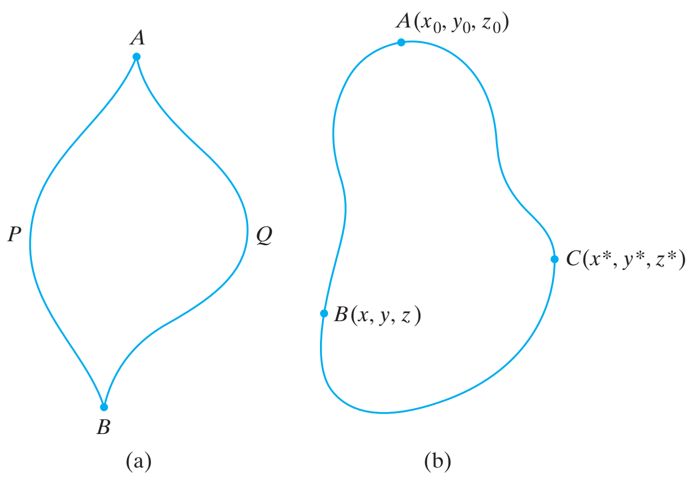


FIGURE 18.14 (a) Two paths joining A to B. (b) A loop containing a fixed point A.

potential field $W_{AB} + W_{BC} + W_{CA} = 0$, so $W_{CA} = -W_{AC} = -\phi(x^*, y^*, z^*)$ and so $W_{BC} = \phi(x^*, y^*, z^*) - \phi(x, y, z)$. This shows that the work done by the force moving between any two points in a potential field is equal to the difference of the potential between the two points.

A gravitational field is due to the presence of matter, so in free space between the matter producing a gravitational force field there can be no sources, and so div $\mathbf{F} = 0$. This means that in a potential field div grad $\phi = 0$ or $\Delta\phi = 0$, so a gravitational potential ϕ is seen to be a solution of the Laplace equation.

the Poisson equation and its connection with the Laplace equation

The linear second order PDE called the **Poisson equation** is

$$\Delta\phi = F(x, y, z), \tag{62}$$

and it is also a PDE of *elliptic* type. The Poisson equation arises in a variety of ways, one of which is in electrostatics when a charge distribution is present in a dielectric medium so that div $\mathbf{E} = \rho/\varepsilon$. If we set $F(x, y, z) = \rho/\varepsilon$, and again introduce an electric potential through $\mathbf{E} = \text{grad}\,\phi$, the equation div $\mathbf{E} = \rho/\varepsilon$ becomes the three-dimensional Poisson equation in (62).

Electromagnetic Waves

Finally, we use Maxwell's equations to show how the wave equation in three space dimensions and time determines electromagnetic wave propagation though space. Returning to the equations in (59), and considering the situation in a dielectric medium where no current can flow so $\mathbf{j} = \mathbf{0}$ and where there is no charge distribution so $\rho = 0$, the equations reduce to

electromagnetic waves in space

$$\text{curl}\,\mathbf{H} = \varepsilon\frac{\partial\mathbf{E}}{\partial t} \quad \text{and} \quad \text{curl}\,\mathbf{E} = -\mu\frac{\partial\mathbf{H}}{\partial t}.$$

Differentiating the first equation with respect to t and substituting for $\partial\mathbf{H}/\partial t$ from the second equation gives

$$-\text{curl curl}\,\mathbf{E} = \varepsilon\mu\,\mathbf{E}_{tt},$$

but curl curl $\mathbf{E} = \text{grad div}\,\mathbf{E} - \Delta\mathbf{E}$, but div $\mathbf{E} = 0$, and so

$$\mathbf{E}_{tt} = (1/\varepsilon\mu)\Delta\mathbf{E}. \tag{63}$$

We have shown that the electric vector $\mathbf{E}$ is a solution of the three-dimensional wave equation.

A similar argument shows that the magnetic vector $\mathbf{H}$ is also a solution of the same three-dimensional wave equation

$$\mathbf{H}_{tt} = (1/\varepsilon\mu)\Delta\mathbf{H}, \tag{64}$$

so that waves involving both the electric and the magnetic vector propagate with the same speed c that is determined by $c^2 = 1/(\varepsilon\mu)$. In free space the speed of propagation c of these electromagnetic waves is the velocity of light.

Summary

Using typical physical examples, the three fundamental types of linear constant coefficient second order PDEs have been derived from first principles. These are the wave equation that is of hyperbolic type, the heat or diffusion equation that is of parabolic type, and the Laplace equation that is of elliptic type. Potential functions and conservative fields were also defined and interpreted in terms of a force acting on a particle moving in the field.

18.6 Classification and Reduction to Standard Form of a Second Order Constant Coefficient Partial Differential Equation for $u(x, y)$

In the previous section the three fundamental types of PDE were derived from typical physical situations, and they were then classed as being of hyperbolic, parabolic, or elliptic type. The purpose of the present section is to explain the basis of this classification where, for simplicity, in the main the discussion will be limited to linear second order partial differential equations whose coefficients are either constants or functions of the independent variables involved.

We have already seen that in the case of two dimensions, examples of these equations involving a function u are as follows:

The one-dimensional wave equation

$$u_{tt} = c^2 u_{xx} \tag{65}$$

the three fundamental types of second order PDE

for the function $u(x, t)$, where x is a space variable, t is the time, and c is a constant.

The one-dimensional heat equation

$$u_t = \kappa^2 u_{xx} \tag{66}$$

for the function $u(x, t)$, where x is a space variable, t is the time, and κ is a constant.

The two-dimensional Laplace equation

$$u_{xx} + u_{yy} = 0 \tag{67}$$

for the function $u(x, y)$, where x and y are both space variables.

These three equations are all special cases of the general linear PDE for an unknown twice differentiable classical solution $u(x, y)$ of the two independent variables x and y, or sometimes t and x, which is defined in some region D and can be written

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Pu_x + Qu_y + Ru = F(x, y), \tag{68}$$

where A, B, C, P, Q, and R are functions of x and y.

In equation (68) the factor 2 multiplying B has been introduced for convenience as it simplifies the calculations that are to follow. The functions $A, B, \ldots, R$ multiplying u and its derivatives are called the **coefficients** of the PDE, and $F(x, y)$ is

called the **nonhomogeneous term**. Equation (68) is called **homogeneous** if $F(x, y)$ is identically zero.

The two-dimensional Laplace equation (67) is an example of a homogeneous constant coefficient PDE that can be derived from (68) by setting $A = C = 1$, $B = 0$, and $F(x, y) = 0$. The corresponding nonhomogeneous equation

$$u_{xx} + u_{yy} = F(x, y) \tag{69}$$

is the **two-dimensional Poisson equation**.

The operations of partial differentiation with respect to x and y are linear when performed on $u(x, y)$, so if two functions $u_1(x, y)$ and $u_2(x, y)$ are solutions of the nonhomogeneous equation (68), it follows that their difference $v(x, y) = u_1(x, y) - u_2(x, y)$ will be a solution of the *homogeneous* equation

$$Av_{xx} + 2Bv_{xy} + Cv_{yy} + Pv_x + Qv_y + Rv = 0. \tag{70}$$

An immediate extension of this result that will be needed later is that if $u_i(x, y)$ with $i = 1, 2, \ldots, k$ are solutions of the homogeneous equation and $c_1, c_2, \ldots, c_k$ are constants, then

$$u(x, y) = \sum_{i=1}^{k} c_i u_i(x, y) \tag{71}$$

is also a solution of the homogeneous equation.

To understand why the three PDEs in (65) to (67) have fundamentally different mathematical properties, it is necessary to examine their mathematical **classification** according to type.

To arrive at the method of classification of second order linear constant coefficient PDEs, we consider the group of second order terms $L[u]$ in (68) given by

$$L[u] = Au_{xx} + 2Bu_{xy} + Cu_{yy}, \tag{72}$$

the quadratic form used to classify a PDE

called the **principal part** of (68), and at some point (x_0, y_0) in a region D where the equation is defined associate with it the *quadratic form*

$$Q(\alpha, \beta) = A(x_0, y_0)\alpha^2 + 2B(x_0, y_0)\alpha\beta + C(x_0, y_0)\beta^2, \tag{73}$$

where α and β are real variables. The differential equation in (68) is then classified according to the following criteria:

classification of PDEs according to type

(a) the PDE is of **hyperbolic** type in D if $B^2 - AC > 0$
(b) the PDE is of **parabolic** type in D if $B^2 - AC = 0$ (74)
(c) the PDE is of **elliptic** type in D if $B^2 - AC < 0$

The expression

$$d = B^2 - AC \tag{75}$$

is called the **discriminant** of the PDE, so it is hyperbolic if $d > 0$, parabolic if $d = 0$, and elliptic if $d < 0$.

When this system of classification is applied to equations (65) to (67), it is seen that the wave equation (65) is of *hyperbolic* type, the heat equation (66) is of *parabolic* type, and the Laplace equation (67) is of *elliptic* type, as is the Poisson equation in (62) because the nonhomogeneous term does not enter into the classification.

This apparently arbitrary classification of the PDEs in (68) is of fundamental importance for the following reasons:

why the classification of PDEs is important

(a) The classification of a PDE is *independent* of the choice of coordinate system used when formulating the equation. Expressed differently, the classification is such that it does not depend on the choice of independent variables. So, for example, if a PDE is of elliptic type when expressed in terms of the cartesian coordinates x and y, it will still be of elliptic type when expressed in terms of any other coordinate system like the cylindrical polar coordinates r, θ, and z.

(b) The nature of an appropriate domain D and the associated auxiliary conditions (initial and/or boundary conditions) that must be imposed on the PDE in order to ensure a unique solution throughout D differ according to the classification.

We will only justify statement (a), as the significance of (b) will become apparent when boundary and initial conditions are considered. Let us make a transformation of the independent coordinate variables x and y to ξ and η in such a way that one point in the domain D in the (x, y)-plane corresponds to one point in the corresponding domain in the (ξ, η)-plane, and conversely (the transformation is one-one between the two domains), by setting

$$\xi = \xi(x, y) \quad \text{and} \quad \eta = \eta(x, y), \tag{76}$$

where the functions ξ and η are assumed to be twice continuously differentiable. The transformation will be one-one if its Jacobian $J(x, y)$ is nonvanishing throughout D, where

$$J(x, y) = \frac{\partial(\xi, \eta)}{\partial(x, y)} = \begin{vmatrix} \xi_x & \xi_y \\ \eta_x & \eta_y \end{vmatrix} \neq 0. \tag{77}$$

Using the rules from the calculus for a change of variables to express the partial derivatives of u with respect to x and y in terms of those with respect to ξ and η, we find that

$$u_x = \xi_x u_\xi + \eta_x u_\eta, \tag{78}$$

so dropping the variable u we obtain the operator relationship

$$\frac{\partial}{\partial x} = \xi_x \frac{\partial}{\partial \xi} + \eta_x \frac{\partial}{\partial \eta} \tag{79}$$

with the corresponding result

$$u_y = \xi_y u_\xi + \eta_y u_\eta \tag{80}$$

and the associated operator relationship

$$\frac{\partial}{\partial y} = \xi_y \frac{\partial}{\partial \xi} + \eta_y \frac{\partial}{\partial \eta}. \tag{81}$$

To find u_{xx} we start from its definition and proceed as follows:

$$u_{xx} = \frac{\partial(u_x)}{\partial x} = \frac{\partial}{\partial x}(\xi_x u_\xi + \eta_x u_\eta) = \xi_{xx} u_\xi + \eta_{xx} u_\eta + \xi_x \frac{\partial(u_\xi)}{\partial x} + \eta_x \frac{\partial(u_\eta)}{\partial x}.$$

Next, replacing the operator $\partial/\partial x$ by the result in (79), simplifying the result, and using the equality of mixed derivatives $u_{\xi\eta} = u_{\eta\xi}$, which is justified because we are considering classical solutions that are continuously twice differentiable, we find

that

$$u_{xx} = \xi_x^2 u_{\xi\xi} + 2\xi_x \eta_x u_{\xi\eta} + \eta_x^2 u_{\eta\eta} + \xi_{xx} u_\xi + \eta_{xx} u_\eta. \tag{82}$$

Similar arguments show that

$$u_{xy} = \xi_x \xi_y u_{\xi\xi} + (\xi_x \eta_y + \xi_y \eta_x) u_{\xi\eta} + \eta_x \eta_y u_{\eta\eta} + \xi_{xy} u_\xi + \eta_{xy} u_\eta \tag{83}$$

and

$$u_{yy} = \xi_y^2 u_{\xi\xi} + 2\xi_y \eta_y u_{\xi\eta} + \eta_y^2 u_{\eta\eta} + \xi_{yy} u_\xi + \eta_{yy} u_\eta. \tag{84}$$

When working with transformations of derivatives, the use of the suffixes x and y with u denoting partial differentiation with respect to x and y is to be understood to imply that u is to be regarded as the *original* function of x and y, but that when the suffixes ξ and η are used it is to be understood that u is then to be regarded as the transformed function $u = u(\xi, \eta)$. The expressions for x and y follow, if the coordinate transformations (76) are solved to obtain $x = \sigma(\xi, \eta)$ and $y = \mu(\xi, \eta)$, which because the transformation is one-one will always enable x and y to be expressed uniquely as functions of ξ and η.

After substituting these results into (68) and collecting terms, we obtain

$$\widetilde{A} u_{\xi\xi} + 2\widetilde{B} u_{\xi\eta} + \widetilde{C} u_{\eta\eta} + \widetilde{P} u_\xi + \widetilde{Q} u_\eta + \widetilde{R} u = \widetilde{F}(\xi, \eta), \tag{85}$$

where

$$\widetilde{A} = A(\xi_x)^2 + 2B\xi_x\xi_y + C(\xi_y)^2 \tag{86}$$
$$\widetilde{B} = A\xi_x\eta_x + B(\xi_x\eta_y + \xi_y\eta_x) + C\xi_y\eta_y \tag{87}$$
$$\widetilde{C} = A(\eta_x)^2 + 2B\eta_x\eta_y + C(\eta_y)^2, \tag{88}$$

with $\widetilde{P}$, $\widetilde{Q}$, and $\widetilde{R}$ defined in similar fashion, and $\widetilde{F}(\xi, \eta) = F(\sigma(\xi, \eta), \mu(\xi, \eta))$.

why a change of variables does not alter the classification of a PDE

A routine calculation establishes the important result that

$$\widetilde{B}^2 - \widetilde{A}\widetilde{C} = (\xi_x \eta_y - \xi_y \eta_x)^2 \,(B^2 - AC) = J^2(x, y)(B^2 - AC).$$

As the Jacobian is nonvanishing and $J^2(x, y)$ is positive, the classification of the equation is seen to be unchanged by the transformation of the independent variables in (76), so statement (a) has been proved.

The transformed PDE in (85) will be simplified if the coordinate transformation can be chosen so that at the point (x_0, y_0):

(a) $\widetilde{A} = \widetilde{C} = 0$, or $\widetilde{A} = -\widetilde{C}$, $\widetilde{B} = 0$, if the PDE is of **hyperbolic** type
(b) $\widetilde{A} = \widetilde{B} = 0$, if the PDE is of **parabolic** type
(c) $\widetilde{A} = \widetilde{C}$ and $\widetilde{B} = 0$, if the PDE is of **elliptic** type.

Clearly this classification depends on the functions $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, and the point (x_0, y_0), though if the original PDE has constant coefficients this classification will be the same for all points in the region D where the PDE is defined.

To see how to accomplish these reductions we again consider the quadratic form $Q(\alpha, \beta)$ in (73) and make the substitutions

$$\alpha = p\xi_x + q\eta_x \quad \text{and} \quad \beta = p\xi_y + q\eta_y,$$

when we find that

$$Q(\alpha, \beta) = \widetilde{A}\lambda^2 + 2\widetilde{B}\lambda\mu + \widetilde{C}\mu^2.$$

This is seen to be of exactly the same algebraic form as the transformation of the principal term $L[u]$ of (68).

So far the functions $\xi(x, y)$ and $\eta(x, y)$ have been arbitrary, so they can now be used to achieve the simplifications in (a), (b), or (c). The **standard forms**, also called **canonical forms**, of the hyperbolic, parabolic, and elliptic PDEs associated with (68) that correspond to cases (a), (b), and (c) are as follows:

Hyperbolic standard forms

$$u_{\xi\eta} = F_1(\xi, \eta, u, u_\xi, u_\eta) \quad \text{or} \quad u_{\xi\xi} = u_{\eta\eta} + F_2(\xi, \eta, u, u_\xi, u_\eta); \tag{89}$$

Parabolic standard form

$$u_{\eta\eta} = G(\xi, \eta, u, u_\xi, u_\eta); \tag{90}$$

Elliptic standard form

$$u_{\xi\xi} + u_{\eta\eta} = H(\xi, \eta, u, u_\xi, u_\eta), \tag{91}$$

where F_1, F_2, G, and H are linear combinations of u, u_ξ, and u_η.

The equivalence of the two different standard forms in the hyperbolic case (89) will be shown later.

Reduction of a Hyperbolic Equation to Standard Form

how to reduce a hyperbolic PDE to standard form

To arrive at the first standard form in (89), ξ and η must be chosen such that $\widetilde{A} = \widetilde{C} = 0$. We see from this that ξ and η must be solutions of the first order PDE

$$A(\varphi_x)^2 + 2B\varphi_x\varphi_y + C(\varphi_y)^2 = 0, \tag{92}$$

which can be factored into the product

$$(A\varphi_x + \{B + \sqrt{B^2 - AC}\}\varphi_y)(A\varphi_x + \{B - \sqrt{B^2 - AC}\}\varphi_y) = 0.$$

Now if φ_1 and φ_2 are solutions of

$$A\varphi_{1x} + (B + \sqrt{B^2 - AC})\varphi_{1y} = 0 \quad \text{and} \quad A\varphi_{2x} + (B - \sqrt{B^2 - AC})\varphi_{2y} = 0, \tag{93}$$

characteristic equations and characteristic curves

they are also solutions of (92). These are called the **characteristic equations** associated with PDE (68), and as the discriminant $d = B^2 - AC > 0$ it follows from Section 18.2 that each defines a family of **characteristic curves** of PDE (68) determined by the solutions of

$$\frac{dy}{dx} = \frac{B + \sqrt{B^2 - AC}}{A} \quad \text{and} \quad \frac{dy}{dx} = \frac{B - \sqrt{B^2 - AC}}{A}. \tag{94}$$

These solutions can be written

$$\varphi_1(x, y) = \text{constant} \quad \text{and} \quad \varphi_2(x, y) = \text{constant}, \tag{95}$$

so we now define the functions ξ and η in (76) as

$$\xi = \varphi_1(x, y) \quad \text{and} \quad \eta = \varphi_2(x, y). \tag{96}$$

With this change of variables (68), and hence (85), reduces to

$$2\widetilde{B}u_{\xi\eta} + \widetilde{P}u_\xi + \widetilde{Q}u_\eta + Ru = \widetilde{F}(\xi, \eta), \tag{97}$$

so

$$u_{\xi\eta} = \frac{1}{2\widetilde{B}}[\widetilde{F}(\xi, \eta) - \widetilde{P}u_\xi - \widetilde{Q}u_\eta - \widetilde{R}u], \tag{98}$$

from which the first result in (89) follows by setting $\widetilde{F}_1(\xi, \eta, u, u_\xi, u_\eta) = [\widetilde{F}(\xi, \eta) - \widetilde{P}u_\xi - \widetilde{Q}u_\eta - \widetilde{R}u]/(2\widetilde{B})$.

The equivalence of the two different standard forms in (90) is established by making the substitution $\xi = X + Y$, $\eta = X - Y$ in $u_{\xi\eta} = F_1(\xi, \eta, u, u_\xi, u_\eta)$. This transforms the equation into $u_{XX} - u_{YY} = F_2(X, Y, u, u_X, u_Y)$, and apart from a change of notation the two results are the same, because F_2 is simply the transformation of F_1.

In the hyperbolic case the discriminant d is positive, so the two families of characteristic curves associated with (68) are two separate families of *real* curves in the (x, y)-plane.

EXAMPLE 18.6 Reduce to standard form

$$u_{xx} + 8u_{xy} + 7u_{yy} + u_x + 2u_y + 3u + y = 0,$$

and find its characteristic equations and curves.

Solution Identifying the PDE with (68) shows $A = 1$, $B = 4$, and $C = 7$, so as the discriminant $d = 4^2 - (1)(7) = 9 > 0$, the equation is hyperbolic. It is unconditionally hyperbolic because the coefficients of the PDE do not depend on position. From (94) the characteristic equations are

$$\frac{dy}{dx} = 1 \quad \text{and} \quad \frac{dy}{dx} = 7.$$

Integrating these equations shows the characteristic curves to be given by the two families of parallel straight lines

$$y = x + \alpha \quad \text{and} \quad y = 7x + \beta,$$

where α and β are arbitrary constants of integration.

Setting $\xi = \alpha = y - x$ and $\eta = \beta = y - 7x$ allows the principal terms in the PDE to be replaced by $2\widetilde{B}u_{\xi\eta}$, and simple calculations establish that $\widetilde{B} = -18$, $u_x = -(u_\xi + 7u_\eta)$, $u_y = u_\xi + u_\eta$, and $y = \frac{1}{6}(7\xi - \eta)$. Substituting for u_x, u_y, and y in the PDE and rearranging terms leads to its being expressed in the standard form

$$u_{\xi\eta} = \frac{1}{36}\left[u_\xi - 5u_\eta + 3u + \frac{1}{6}(7\xi - \eta)\right]. \quad \blacksquare$$

Reduction of a Parabolic Equation to Standard Form

The standard form in (90) arises when the discriminant $d = B^2 - AC = 0$, in which case the two characteristic equations in (94) coincide and so determine only one family of characteristic curves given by

how to reduce a parabolic PDE to standard form

$$\frac{dy}{dx} = \frac{B}{A} \quad \text{with the characteristics } y = (B/A)x + \alpha, \tag{99}$$

where α is an arbitrary constant of integration.

The required reduction is accomplished by equating ξ and α and choosing for η *any* function of x and y that is independent of ξ, so in general we can set $\eta = x$. Then with the change of variables

$$\xi = y - (B/A)x, \quad \eta = x, \tag{100}$$

the principal terms of PDE (68) can be replaced by $Au_{\eta\eta}$, so that (85) becomes

$$\widetilde{A}u_{\eta\eta} + \widetilde{P}u_{\xi} + \widetilde{Q}u_{\eta} + \widetilde{R}u = \widetilde{F}(\xi, \eta),$$

and so

$$u_{\eta\eta} = \frac{1}{\widetilde{A}}\,[\widetilde{F}(\xi, \eta) - \widetilde{P}u_{\xi} - \widetilde{Q}u_{\eta} - \widetilde{R}u], \tag{101}$$

from which (90) follows by setting $G(\xi, \eta, u, u_{\xi}, u_{\eta}) = (1/\widetilde{A})[\widetilde{F}(\xi, \eta) - \widetilde{P}u_{\xi} - \widetilde{Q}u_{\eta} - \widetilde{R}u]$.

EXAMPLE 18.7

Reduce to standard form

$$u_{xx} + 4u_{xy} + 4u_{yy} + u_x + 3x = 0.$$

Solution Here $A = 1$, $B = 2$, and $C = 4$, so the discriminant $d = B^2 - AC = 0$, showing that the PDE is unconditionally parabolic. In this case the transformation $\xi = y - (B/A)x$ and $\eta = x$ becomes $\xi = y - 2x$, $\eta = x$, and this change of variables allows the principal terms to be replaced by $Au_{\eta\eta}$, so as $u_x = -2u_{\xi} + u_{\eta}$ and $x = \eta$, substitution into the PDE leads to the required reduction to standard form

$$u_{\eta\eta} = 2u_{\xi} - u_{\eta} - 3\eta. \qquad \blacksquare$$

Reduction of an Elliptic Equation to Standard Form

When PDE (68) is elliptic, its discriminant $d = B^2 - AC < 0$, so the right-hand sides of the characteristic equations in (94) become complex, showing that an elliptic PDE has no real characteristic curves. However, in the elliptic case the transformation

how to reduce an elliptic PDE to standard form

$$\xi = \frac{Ay - Bx}{\sqrt{AC - B^2}}, \quad \eta = x, \tag{102}$$

reduces (68) to

$$A(u_{\xi\xi} + u_{\eta\eta}) + \widetilde{P}u_{\xi} + \widetilde{Q}u_{\eta} + \widetilde{R}u = \widetilde{F}(\xi, \eta), \tag{103}$$

so as $A \neq 0$ an elliptic equation can always be written in the standard form

$$u_{\xi\xi} + u_{\eta\eta} = \frac{1}{A}[\widetilde{F}(\xi,\eta) - \widetilde{P}u_\xi - \widetilde{Q}u_\eta - \widetilde{R}u], \tag{104}$$

that is, of the form in (91) with $H(\xi,\eta,u,u_\xi,u_\eta) = (1/A)[\widetilde{F}(\xi,\eta) - \widetilde{P}u_\xi - \widetilde{Q}u_\eta - \widetilde{R}u]$.

EXAMPLE 18.8 Reduce to standard form

$$5u_{xx} - 2u_{xy} + 2u_{yy} + 2u_y + 4y = 0.$$

Solution Here $A = 5$, $B = -1$, and $C = 2$, so the discriminant $d = B^2 - AC = -9$, showing that the PDE is unconditionally elliptic. From (102) the transformation to be used is $\xi = \frac{1}{3}(5y - x)$ and $\eta = x$, and when this change of variables has been made the principal terms can be replaced by $A(u_{\xi\xi} + u_{\eta\eta})$, so substituting into (103) and using the results $u_y = \frac{5}{3}u_\eta$ and $y = \frac{1}{5}(3\xi - \eta)$ gives the required reduction

$$u_{\xi\xi} + u_{\eta\eta} = \frac{1}{75}[12\eta - 36\xi - 50u_\xi].$$ ■

EXAMPLE 18.9 Classify and reduce to standard form the PDE

$$u_{xx} + yu_{yy} + \frac{1}{2}u_y + 4yu_x = 0.$$

Solution This is now a variable coefficient PDE with $A = 1$, $B = 0$, and $C = y$, so the discriminant $d = B^2 - AC = -y$. This shows the equation to be elliptic when $y > 0$, hyperbolic when $y < 0$, and degenerately parabolic on the x-axis.

Elliptic Case $y > 0$

The characteristic equations become

$$\frac{dy}{dx} = -\sqrt{-y} \quad \text{or} \quad \frac{dy}{dx} = -i\sqrt{y}, \quad \text{and} \quad \frac{dy}{dx} = \sqrt{-y} \quad \text{or} \quad \frac{dy}{dx} = i\sqrt{y}.$$

Integrating these complex characteristic equations gives

$$2\sqrt{y} = -ix + \xi - i\eta \quad \text{and} \quad 2\sqrt{y} = ix + \xi + i\eta,$$

and solving for ξ and η we find that $\xi = 2\sqrt{y}$ and $\eta = -x$.

Substituting into (78), (80), (82), and (84) gives

$$u_x = -u_\eta, \quad u_y = \frac{1}{\sqrt{y}}u_\xi, \quad u_{xx} = u_{\eta\eta}, \quad \text{and} \quad u_{yy} = \frac{1}{y}u_{\xi\xi} - \frac{1}{2y^{3/2}}.$$

Using these results to transform the original PDE gives the standard form

$$u_{\xi\xi} + u_{\eta\eta} = \xi^2 u_\eta - \frac{1}{2}(1 - 2/\xi)u_\xi.$$

Hyperbolic Case $y < 0$

The characteristic equations become

$$\frac{dy}{dx} = -\sqrt{-y} \quad \text{and} \quad \frac{dy}{dx} = \sqrt{-y},$$

with the respective solutions

$$-2\sqrt{-y} = -x + \xi \quad \text{and} \quad -2\sqrt{-y} = x + \eta,$$

so

$$\xi = x - 2\sqrt{-y} \quad \text{and} \quad \eta = -x - 2\sqrt{-y}.$$

Substituting into (78), (80), (82), and (84) gives

$$u_x = u_\xi - u_\eta, \quad u_y = (1/\sqrt{-y})(u_\xi + u_\eta), \quad u_{xx} = u_{\xi\xi} - 2u_{\xi\eta} + u_{\eta\eta},$$

$$u_{yy} = -(1/y)u_{\xi\xi} - (2/y)u_{\xi\eta} - (1/y)u_{\eta\eta} + \frac{1}{2}(-y)^{-3/2}(u_\xi + u_\eta).$$

When these are substituted into the original PDE it becomes

$$u_{\xi\eta} = \frac{1}{16}(\xi + \eta)^2(u_\eta - u_\xi) - \left(\frac{1}{\xi + \mu}\right)(u_\xi + u_\eta).$$

This classification of PDEs can be extended to equations with n independent variables by using the property of orthogonal matrices, which were introduced in Chapter 4. Let the second order constant coefficient PDE for an unknown function $u(x_1, x_2, \ldots, x_n)$ in the n independent variables $x_1, x_2, \ldots, x_n$ be written

classification of PDEs in n independent variables

$$\sum_{i,j=1}^{n} a_{ij}u_{x_i x_j} + \sum_{i=1}^{n} b_i u_{x_i} + cu = F(x_1, x_2, \ldots, x_n), \tag{105}$$

where the coefficients a_{ij}, b_i, and c are real constants and F is a real function of its n arguments. Then, because of the equivalence of mixed partial derivatives, it is always possible to assume the a_{ij} to be symmetric and to write $a_{ij} = a_{ji}$.

We now define an n element column vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ involving the independent variables, and make a linear transformation of $\mathbf{x}$ to a new set of variables $\xi_1, \xi_2, \ldots, \xi_n$ that can be written as the column vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_n]^{\mathrm{T}}$. The linear transformation can be expressed in terms of an $n \times n$ matrix $\mathbf{B} = [b_{ij}]$ with real elements by writing

$$\boldsymbol{\xi} = \mathbf{B}\mathbf{x}. \tag{106}$$

As with second order PDEs in two independent variables, the classification of the second order PDE (105) is determined by the way in which

$$L[u] = \sum_{i,j=1}^{n} a_{ij}u_{x_i x_j} \tag{107}$$

transforms into a standard form that is free from mixed derivatives, so we need only consider the effect of this linear transformation on its principal part $L[u]$, the result of which can be written

$$L[u] = \sum_{i,j=1}^{n} a_{ij}u_{x_i x_j} = \sum_{p,q=1}^{n}\left(\sum_{i,j=1}^{n} b_{pi}a_{ij}b_{qj}\right)u_{\xi_p \xi_q}. \tag{108}$$

In matrix form this transformation of the leading terms is seen to have the coefficient matrix $\mathbf{BAB}^{\mathrm{T}}$. As $\mathbf{A}$ is symmetric, its eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ will all

be *real*, and it follows from Theorem 4.10 that an orthogonal matrix $\mathbf{Q}$ can always be associated with $\mathbf{A}$ in such a way that $\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ as the elements along its leading diagonal. Consequently, if we set $\mathbf{B} = \mathbf{Q}^{\mathrm{T}}$, and

$$\mathbf{Q} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

the principal terms of PDE (105) become

$$L(u) = \lambda_1 u_{\xi_1\xi_1} + \lambda_2 u_{\xi_2\xi_2} + \cdots + \lambda_n u_{\xi_n\xi_n}. \tag{109}$$

A simple scaling of the variables $\xi_1, \xi_2, \ldots, \xi_n$ will always reduce the principal terms in $L[u]$ to the form

$$L(u) = \varepsilon_1 u_{\xi_1\xi_1} + \varepsilon_2 u_{\xi_2\xi_2} + \cdots + \varepsilon_n u_{\xi_n\xi_n}, \tag{110}$$

where ε_i is $+1$ when $\lambda_i > 0$, -1 when $\lambda_i < 0$, and 0 when $\lambda_i = 0$.

The classification of PDE (105) involves a generalization of the case of two independent variables to n independent variables as follows:

classification according to type of PDEs in *n* independent variables

(a) PDE (105) is of **hyperbolic type** if none of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ vanishes and only one eigenvalue has a sign opposite to that of the remaining $n-1$ eigenvalues. So, if the eigenvalues are ordered such that $\lambda_1 > 0$, after scaling the independent variables $\xi_1, \xi_2, \ldots, \xi_n$ a hyperbolic PDE of type (105) will have the **standard form**

$$u_{\xi_1\xi_1} = u_{\xi_2\xi_2} + u_{\xi_3\xi_3} \cdots + u_{\xi_n\xi_n} + F(\xi_1, \ldots, \xi_n, u, u_{\xi_1}, \ldots, u_{\xi_n}), \tag{111}$$

where F is a linear combination of $u, u_{\xi_1}, \ldots, u_{\xi_n}$.

(b) PDE (105) is of **parabolic type** if one of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ vanishes and the remaining $n-1$ eigenvalues are all of the same sign. So, if the eigenvalues are ordered so that $\lambda_1 = 0$, after scaling the independent variables $\xi_1, \xi_2, \ldots, \xi_n$ a parabolic PDE of type (105) will have the **standard form**

$$u_{\xi_2\xi_2} + u_{\xi_3\xi_3} \cdots + u_{\xi_n\xi_n} = G(\xi_1, \ldots, \xi_n, u, u_{\xi_1}, \ldots, u_{\xi_n}), \tag{112}$$

where G is a linear combination of $u, u_{\xi_1}, \ldots, u_{\xi_n}$.

(c) PDE (105) is of **elliptic type** if none of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ vanishes and all have the same sign that may be either positive or negative. So after scaling the independent variables $\xi_1, \xi_2, \ldots, \xi_n$ an elliptic PDE of type (105) will have the **standard form**

$$u_{\xi_1\xi_1} + u_{\xi_2\xi_2} + u_{\xi_3\xi_3} \cdots + u_{\xi_n\xi_n} = H(\xi_1, \ldots, \xi_n, u, u_{\xi_1}, \ldots, u_{\xi_n}), \tag{113}$$

where H is a linear combination of $u, u_{\xi_1}, \ldots, u_{\xi_n}$.

EXAMPLE 18.10 Classify the PDE

$$4u_{x_1x_1} + 4u_{x_2x_2} + u_{x_3x_3} - 2u_{x_1x_2} = 0,$$

and find the form to which it is reduced by an orthogonal transformation that converts its coefficient matrix to a diagonal matrix.

Solution Because of the equality of mixed derivatives, the matrix form of the PDE can be written $\mathbf{AU} = \mathbf{0}$, where

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{U} = \begin{bmatrix} u_{x_1x_1} \\ u_{x_2x_2} \\ u_{x_3x_3} \end{bmatrix}.$$

The eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1, \lambda_2 = 5$, and $\lambda_3 = 3$, so from (a) just shown, the PDE is seen to be of *elliptic* type. As the PDE only contains principal terms, an orthogonal transformation that transforms $\mathbf{A}$ into a diagonal matrix will transform the PDE into

$$u_{\xi_1\xi_1} + 5u_{\xi_2\xi_2} + 3u_{\xi_3\xi_3} = 0.$$

The actual change of variables from x_1, x_2, and x_3 to ξ_1, ξ_2, and ξ_3 necessary to accomplish this was shown in Example 4.18 to be given by $\xi = \mathbf{Qx}$, where $\mathbf{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ and $\xi = [\xi_1, \xi_2, \xi_3]^{\mathrm{T}}$, with the orthogonal matrix $\mathbf{Q}$ and the diagonal matrix $\mathbf{D}$ given by

$$\mathbf{Q} = \begin{bmatrix} 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

So the necessary change of variables determined by $\xi = \mathbf{Qx}$ becomes

$$\xi_1 = -\frac{1}{\sqrt{2}}x_2 + \frac{1}{\sqrt{2}}x_3, \quad \xi_2 = \frac{1}{\sqrt{2}}x_2 + \frac{1}{\sqrt{2}}x_3 \quad \text{and} \quad \xi_3 = x_1.$$

If necessary, the PDE can be further simplified by scaling the variables ξ_1, ξ_2, ξ_3 to arrive at the new variables ζ_1, ζ_2, and ζ_3 by writing

$$\zeta_1 = \xi_1, \quad \zeta_2 = \frac{1}{\sqrt{5}}\xi_2 \quad \text{and} \quad \zeta_3 = \frac{1}{\sqrt{3}}\xi_3,$$

because then the PDE reduces to the standard form

$$u_{\zeta_1\zeta_1} + u_{\zeta_2\zeta_2} + u_{\zeta_3\zeta_3} = 0,$$

which is Laplace's equation in three independent variables. ■

For more information on the classification of PDEs, see references [7.6] and [7.19].

Summary

Linear second order PDEs in two independent variables have been classified and shown to belong to one of three distinct types, namely, PDEs of hyperbolic, parabolic, and elliptic type. Changes of variable were introduced that simplified the structure of each type of equation by reducing it to one of three standard forms. In each case, the method of reduction to standard form was illustrated by an example, and the classification was then extended to linear second order PDEs in n independent variables.

EXERCISES 18.6

In Exercises 1 through 6 classify the given PDE.

1. $4u_{xx} - 6u_{xy} + 3u_{yy} + 2u_x + 6 = 0.$
2. $u_{xx} + 8u_{xy} - 2u_{yy} + u_x + 3u_y + 2u - 3 = 0.$
3. $2u_{xx} - 2u_{xy} + u_{yy} + 4u_x + 2u + 1 = 0.$
4. $4u_{xx} - 4u_{xy} + u_{yy} + 6u_x - u_y + (1 + x)u + 2 = 0.$
5. $3u_{xx} + 6u_{xy} + 3u_{yy} + (1 + \sin x)u = 0.$
6. $2u_{xx} + 2u_{xy} - u_{yy} + 3u_y + u + 5 = 0.$

In Exercises 7 through 12 classify and reduce to standard form the given PDE.

7. $u_{xx} - 2u_{xy} + 5u_{yy} + 3u_x + 1 = 0.$
8. $4u_{xx} + 4u_{xy} + u_{yy} + 4u_y + u = 0.$
9. $u_{xx} - 10u_{xy} + 9u_{yy} + u_x = 0.$
10. $u_{xx} - 4u_{xy} - 5u_{yy} + 3u_y + u + 4 = 0.$
11. $u_{xx} + 6u_{xy} + 9u_{yy} - u + 5 = 0.$
12. $2u_{xx} - 6u_{xy} + 5u_{yy} + 4u_x + u_y - 2 = 0.$

In Exercises 13 through 16 classify the PDE, and by using a suitable orthogonal matrix **Q** followed, if necessary, by a scaling of the independent variables, reduce it to standard form.

13.* $5u_{x_1x_1} + 2u_{x_2x_2} + 8u_{x_2x_3} + 2u_{x_2} + 4u + 1 = 0.$
14.* $2u_{x_2x_2} - 4u_{x_1x_3} + u_{x_3} + 1 = 0.$
15.* $3u_{x_1x_1} + 2u_{x_2x_2} - 2u_{x_2x_3} + 2u_{x_3x_3} + 4u - 7 = 0.$
16.* $u_{x_1x_1} + 2u_{x_2x_3} + u_{x_2} + 5u + 2 = 0.$

18.7 Boundary Conditions and Initial Conditions

The PDEs derived in Section 18.5, and classified in Section 18.6, are special cases of the general linear PDE for an unknown function $u(x, y)$ of the two independent variables x and y

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Pu_x + Qu_y + Ru = F(x, y), \tag{114}$$

though sometimes with y replaced by t.

Physical problems whose solution is governed by a PDE of this type are formulated in some region D of the (x, y)-plane on the boundary Γ of which suitable auxiliary conditions, called **boundary conditions**, are imposed that serve to identify a particular problem. The most important types of boundary conditions are as follows:

(a) The specification of the functional form to be taken by the solution $u(x, y)$ on the boundary Γ, by requiring that

$$u(x, y) = \Phi(x, y) \quad \text{for } (x, y) \text{ on } \Gamma, \tag{115}$$

Dirichlet boundary condition

where $\Phi(x, y)$ is a given function. A boundary condition of this type is called a **Dirichlet condition**.

(b) The specification of the functional form to be taken by the derivative of the solution $u(x, y)$ normal to the boundary Γ, by requiring that

$$\frac{\partial u}{\partial n}(x, y) = \Psi(x, y) \quad \text{for } (x, y) \text{ on } \Gamma, \tag{116}$$

Neumann boundary condition

where $\Psi(x, y)$ is a given function and $\partial/\partial n$ is the directional derivative normal to the boundary Γ. A boundary condition of this type is called a **Neumann condition**.

(c) The specification of the functional form to be taken by a linear combination of a Dirichlet condition and a Neumann condition by the solution $u(x, y)$ on the boundary Γ, by requiring that

$$a(x, y)u(x, y) + b\frac{\partial u}{\partial n}(x, y) = c(x, y) \quad \text{for } (x, y) \text{ on } \Gamma, \tag{117}$$

mixed or Robin boundary condition

where $a(x, y)$, $b(x, y)$, and $c(x, y)$ are given functions. A boundary condition of this type is called a **mixed condition**, and sometimes either a **Robin condition** or a **boundary condition of the third kind**. When $c(x, y) = 0$, this condition is called a **homogeneous mixed condition**.

(d) The specification on Γ of the functional form to be taken by both the solution $u(x, y)$ and its derivative normal to the boundary, by requiring that

$$u(x, y) = \Phi(x, y) \quad \text{and} \quad \frac{\partial u}{\partial n}(x, y) = \Psi(x, y) \quad \text{for } (x, y) \text{ on } \Gamma, \tag{118}$$

Cauchy conditions

where $\Phi(x, y)$ and $\Psi(x, y)$ are given functions and $\partial/\partial n$ is the directional derivative normal to the boundary Γ. Boundary conditions of this type are called **Cauchy conditions** for a second order PDE.

When the solution u is a function of a space variable x and the time t, and Cauchy conditions are specified when $t = 0$, so that Γ becomes the x-axis and

$$u(x, 0) = \Phi(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(x, 0) = \Psi(x), \tag{119}$$

initial conditions

the Cauchy conditions are usually called **initial conditions** for a second order PDE.

The types of boundary condition that can be imposed on PDE (114) depend on its classification and the nature of the region D that is involved. Some typical examples of boundary conditions and their associated regions for PDEs of hyperbolic, parabolic, and elliptic type were seen in Section 18.5 when the three types of equation were derived from physical problems.

open and closed regions

A region D is classified as being **closed** when it is enclosed by a boundary and every point on the boundary belongs to D, and as being **open** when either the region D extends to infinity or, although D is contained within a boundary, not all of the points of the boundary belong to D. Typical closed regions are the rectangle $a \le x \le b$, $c \le y \le d$, and the annular region $R_1 \le r \le R_2$ centered on the origin. Examples of open regions are the semi-infinite strip $a \le x \le b$, $y \ge 0$, where the boundary points on three sides of the strip belong to the region but there is no upper boundary because y extends to infinity, and the annular region $R_1 < r \le R_2$, where the points on the outer rim of the annulus belong to the region but the points on the inner rim do not.

well-posed and improperly posed problems

When the boundary conditions and the region D are such that a unique solution exists, and small changes in the boundary conditions only produce small changes in the solution, the boundary value problem is said to be **well posed**, and the solution is said to be **stable**. If, however, the boundary conditions and/or region are such that although a solution exists, a small change in the boundary conditions causes a large change in the solution, the boundary value problem is said to be **improperly**

posed and the solution is then said to be **unstable**. In what follows our concern will only be with well-posed problems.

Listed next are the most frequently occurring combinations of boundary conditions and regions that lead to properly posed problems for hyperbolic, parabolic, and elliptic PDEs.

appropriate conditions and regions for the three types of PDE

Type of PDE	Conditions	Type of Region
Hyperbolic	Cauchy conditions	Open
Parabolic	Dirichlet, Neumann, or mixed	Open
Elliptic	Dirichlet, Neumann, or mixed	Closed

The effect of imposing inappropriate boundary conditions on a PDE can lead to one of the following situations: (a) no solution exists, (b) a solution exists, but it is either trivial (identically zero) or not unique, and (c) a solution exists, but it is not stable.

To demonstrate that appropriateness of the preceding conditions, by way of example we prove that the Dirichlet problem for the Laplace equation in a closed region is a properly posed problem. To do this we will make use of Theorem 14.17, which showed that a harmonic function defined in a closed region D with boundary Γ must attain its greatest and least values on the boundary Γ. A trivial corollary of this theorem that will be needed, and that is almost immediately obvious, is that if $u(x, y)$ is harmonic in D and is equal to the constant k on the boundary of Γ of D, then $u(x, y) = k$ throughout D.

Let us use this theorem to prove the uniqueness of a function u that is harmonic in D and satisfies a Dirichlet condition $u|_\Gamma = f(s)$ on Γ, where the parameter s can be taken to be the arc length measured around Γ from some fixed point on the boundary. Suppose, if possible, that this Dirichlet problem has two *different* solutions u and v that satisfy the *same* Dirichlet condition, and set $w = u - v$. Then because the Laplace equation is linear, w is also a solution of Laplace's equation, and on Γ it satisfies the homogeneous boundary condition $w|_\Gamma = 0$. Using the corollary of the maximum–minimum theorem mentioned earlier, it follows at once that $w \equiv 0$ throughout D, and so $u \equiv v$, and the uniqueness of the solution has been established.

As a further demonstration of the appropriateness of Dirichlet conditions for the Laplace equation, we now prove that small changes in the boundary conditions produce small changes in the solution, because this shows the continuous dependence of the solution on the boundary data (Dirichlet condition). Let u_1 and u_2 be solutions of two different Dirichlet problems for the Laplace equation in a closed region D with boundary Γ on which u_1 satisfies the Dirichlet condition $u_1|_\Gamma = f_1(s)$ and u_2 satisfies the Dirichlet condition $u_2|_\Gamma = f_2(s)$, where s is defined as before and $|f_1(s) - f_2(s)| < \varepsilon$ on Γ, with $\varepsilon > 0$ arbitrarily small. The difference $u_1 - u_2$ is also harmonic in D, so the condition on $f_1 - f_2$ is equivalent to $-\varepsilon < u_1 - u_2 < \varepsilon$ on Γ. It follows directly from the maximum–minimum theorem that throughout D we must have $-\varepsilon < u_1 - u_2 < \varepsilon$, so $|u_1 - u_2| < \varepsilon$. This has established that when the Dirichlet data is only changed by a small amount, the same is true of the solution, so the continuous dependence of the solution on the Dirichlet data has been established. This result, combined with the uniqueness of the solution, shows that the Dirichlet problem for the Laplace equation is well posed.

Summary

The main types of boundary condition suitable for second order PDEs were described, and the notion of a well-posed problem was introduced. Open and closed regions were defined and a short table was given listing suitable combinations of boundary condition and region according to the type of PDE involved.

18.8 Waves and the One-Dimensional Wave Equation

The general solution of the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \tag{120}$$

has a useful interpretation in terms disturbances that move with speed c in opposite directions along the x-axis.

It is known from Section 18.6 that the characteristic equations for the wave equation are

$$\frac{dx}{dt} = c \quad \text{and} \quad \frac{dx}{dt} = -c.$$

Integrating the first of these equations to find the characteristic through the point $(\xi, 0)$ on the x-axis (the initial line) gives

$$\xi = x - ct, \tag{121}$$

and integrating the second equation to find the characteristic through the point $(\eta, 0)$ on the x-axis gives

$$\eta = x + ct. \tag{122}$$

Changing the independent variables in (120) from x and t to ξ and η reduces it to the standard form $u_{\xi\eta} = 0$. Integrating this result partially with respect to η, while regarding ξ as a constant in order to reverse the process of partial differentiation, gives $u_\xi = f(\xi)$, where F is an arbitrary differentiable function of ξ. Next, integrating this result partially with respect to ξ, where η is now regarded as a constant, leads to the result $u(\xi, \eta) = f(\xi) + g(\eta)$, where $f(\xi) = \int F(\xi) d\xi$ and g is an arbitrary differentiable function of η. Notice that as $F(\xi)$ is an arbitrary function, so also is $f(\xi)$. Finally, if we revert to the original variables x and t, the general solution of (120) becomes

general solution of wave equation as the sum of two waves moving in opposite directions

$$u(x, t) = f(x - ct) + g(x + ct). \tag{123}$$

The function $f(x - ct)$ will be constant along the characteristic $x - ct =$ constant, so considering all possible characteristics of this type the term $f(x - ct)$, as in Section 18.3, in (123) is seen to transport the initial shape of the function f to the right along the x-axis with constant speed c without change of either shape or scale. Similarly, the term $g(x + ct)$ will transport the initial shape of the function g to the left along the x-axis, also with the constant speed c and without change of shape or scale.

Disturbances that propagate through space at a finite speed as time increases are called **waves**, so the general solution in (123) represents two traveling waves moving in opposite directions, each with the constant speed c. The initial disturbances $f(x)$ and $g(x)$ are called the **wave profiles**, so the shape of each wave profile in (123) is preserved as it propagates.

waves and wave profiles

The interpretation of the general solution (123) of the wave equation in (120) is now clear, because it shows that an initial disturbance $u(x, 0)$ is resolved into two traveling waves, one moving to the right and the other to the left, each with the same speed c and without change of shape or scale. The general solution also shows that the disturbance (wave) propagated by the wave equation at any time t is the *sum* of the disturbances caused by the traveling waves as they move to the left and right, so that (123) describes the *interaction* of the two wave profiles. This very important property of the wave equation is due to its *linearity*, which allows the sum of any two solutions to be another solution.

To make effective use of this result when seeking the solution of a Cauchy problem for the wave equation, it is necessary to know how the initial disturbance $u(x, 0)$ is resolved into the functions $f(x)$ and $g(x)$. We will see how to find $f(x)$ and $g(x)$ in terms of the Cauchy conditions when the D'Alembert solution of the one-dimensional wave equation is derived in the next section.

In Section 18.5 the wave equation was derived under the assumption that $u(x, t)$ is a continuous and twice differentiable function of its arguments. We will now use result (123) to show how these conditions can be relaxed to allow for initial wave profiles that have a discontinuity in their derivative, or even a finite jump discontinuity in the functions themselves.

Suppose that the wave profile $f(x)$ has a discontinuity either in its derivative or in the function itself at some point $x = x^*$. The characteristic through $x = x^*$ does not depend on the solution, so the propagating wave profile can be separated into two distinct parts, one to the left of the characteristic $x - ct = x^*$, and the other to the right.

how to deal with wave profiles with discontinuities

The characteristics to the immediate left and right of $x - ct = x^*$ are both parallel to it. This means that the wave profile to the left of this characteristic will propagate in the manner just described, independently of the wave to the right, but bounded on the right by $x - ct = x^*$. Similarly, the wave profile to the right of this characteristic will propagate in the manner described, independently of the solution to the left, but bounded on the left by $x - ct = x^*$. As the solutions to the immediate left and right move along the same characteristic $x - ct = x^*$, any initial discontinuity in f will be propagated along this characteristic without change.

The same result is true for the initial wave profile $g(x)$, so the interpretation of the general solution (123) as the sum of disturbances due to the two wave profiles propagating to the right and left remains valid even when a discontinuity in the derivative or in the initial disturbance $u(x, 0)$ itself is present. This generalizes the concept of a solution of the wave equation, because it permits initial disturbances with discontinuities in either a derivative or the function itself. This situation is quite different from the quasilinear case considered in Section 18.4, because there a discontinuity in the solution was seen to propagate at the shock speed, which is quite different from that of the adjacent characteristic speeds.

We now use this generalization to examine the resolution of an initial disturbance that is localized in a finite part $a < x < b$ of the x-axis, and zero outside it. The purpose of this is to make clear how the two wave profiles interact until, after a suitable lapse of time, they move clear of one another, after which all interaction